

THIS WEEK

EDITORIALS

WORLD VIEW Halt flawed, biased and poor-quality animal studies **p.511**

SI UNITS Constant approach brings new kilogram and ampere closer **p.512**



IAEA Member states give nuclear watchdog body sharper teeth **p.514**

One up, one down

Sending satellites to monitor the atmosphere is fundamental to weather and climate research. So why is the United States making such a meal of it?

Late next month, barely four weeks after the remains of its dead upper-atmosphere research satellite rained into the Pacific Ocean, the United States is scheduled to send a new mission into polar orbit. A joint project of NASA and the National Oceanic and Atmospheric Administration (NOAA), the spacecraft is to provide basic information for meteorologists. It will also ensure a continuing flow of data to climate researchers, as they have come to expect.

There should be nothing remarkable about this. But the truth is that the scheduled 25 October launch of the awkwardly named NPOESS Preparatory Project (NPP) represents a landmark achievement for the United States' long-troubled polar satellite programme. As recent discussions on Capitol Hill have made apparent, it remains utterly unclear what will follow the launch, and when.

This odd situation has its roots in the 1994 decision by the administration of president Bill Clinton to merge separate defence and civilian programmes into a single National Polar-orbiting Operational Environmental Satellite System (NPOESS).

What seemed a logical decision instead saw the programme spiral into a bureaucratic nightmare involving NOAA, NASA and the Department of Defense. In the end, it fell to President Barack Obama to dismantle the programme entirely and start again. More than a year later, NOAA and NASA are restoring order to the newly restructured Joint Polar Satellite System. (The same cannot be said for a separate component retained by the defence department, which remains in limbo.)

The challenge now, of course, is money. Overall programme costs have more than doubled the original US\$7-billion estimate, with NOAA alone predicting costs of nearly \$12 billion to 2024. Even with this, expected goals have been scaled back and delayed.

With funding flat for the programme in 2011, NOAA now wants \$1.1 billion in fiscal year 2012, an increase of \$688 million, to get the next polar satellite ready for launch in late 2016 or early 2017. And if Congress fails to provide — well, the agency threatens, Americans will have to suffer the consequences of a significantly degraded weather-forecast system. Citing storms such as Hurricane Irene, which hammered the US East Coast in August, NOAA assistant administrator Kathryn Sullivan says that these satellites provide “critical environmental intelligence”.

All of this leaves lawmakers on Capitol Hill rightly confused. Democrats and Republicans alike support the programme, yet are understandably wary and angry about NOAA's ongoing failure to provide a complete rundown of the budget going forward. House and Senate appropriations committees did their best to find the money, but that doesn't mean much in a volatile year such as this one. As Paul Broun, the Georgia Republican who chairs the House Science Subcommittee on Investigations and Oversight, warned NOAA last week, Congress might not pass any funding bills at all, and instead might maintain current spending levels into 2012 and perhaps 2013. And on top of

that, a high-profile congressional panel is busy looking for further savings to reduce the US debt.

NOAA deserves credit for setting the satellite programme back on an even keel, but the agency will need to provide more than dire warnings of disaster and human suffering if the requested funding doesn't come through. Nor does it help when the agency fails to meet legitimate demands for information from lawmakers — and the press. In fact, on the evidence of *Nature's* rebuffed enquiries, it seems harder

“NOAA should do as its supporters in Congress suggest, and prepare a proper budget for the programme.”

to get information out of NOAA than from the Department of Energy's nuclear weapons “programme. NOAA should do as its supporters in Congress suggest, and prepare a proper budget for the programme, including contingency plans in case funds fail to materialize.

Some options exist. Lawmakers are already thinking about ways in which NOAA can charge for some of the services it provides for other agencies and institutions. This might be an effective way to spread some of the pain, but if the agency moves in this direction it should be for value-added services only. The raw satellite data themselves must remain freely available, both at home and abroad.

In the meantime, fingers are being kept firmly crossed that the NPP satellite makes it safely into space and performs as well as its predecessors. Beyond that, well, NASA officials say they have no reason to believe that the satellite's sensors are going to suddenly shut down at the end of their mission life, as NOAA has implied in its push for a successor. Indeed, many satellites run for years beyond their scheduled lifetime. It would be foolhardy to run a programme on this basis, but it may yet come to that. ■

Blackened names

Canada should make public the identities of researchers who commit misconduct.

When government agencies redact material from released records, there are inevitably questions. That was certainly the case on 19 September, when Margaret Munro, a reporter for Postmedia News in Canada, revealed details of several seemingly egregious cases of misconduct that she had obtained under Access to Information laws.

The heavily censored documents, released by the Natural Sciences and Engineering Research Council of Canada (NSERC), the country's

science-funding agency, described a case in which a researcher had faked data, then moved to another institution that may have known nothing of the episode. A second, even more remarkable, case involved a scientist who padded his CV with papers that did not exist anywhere in the published literature.

But the names of the wrong-doers were blacked out, as were the citations of retracted papers and the details of the research fields affected. In some cases, information on the misconduct findings and the nature of the wrong-doing was also redacted. One document retained sentences of a letter in which someone — presumably a misconduct perpetrator — offered abject apologies and argued that cutting off his research funding would harm his students, but blacked out the name and numbers of people involved. In another, a black rectangle obscured a list of six problematic journal articles, some of which are apparently fictitious.

Canada's practices take privacy concerns too far. There is an argument for confidentiality while an investigation is going on, or even after it has been resolved, if allegations turn out to be frivolous or malicious. In the United States, many agencies withhold the names of those accused of misconduct. Some officials argue that cases of honest error should also remain confidential, to protect the careers of scientists exonerated of misconduct. But these arguments cease to be relevant once misconduct has been found, as the US government's Office of Research Integrity acknowledges when it makes public the names of those found guilty of misconduct.

Taxpayers have a right to know about instances in which their money has been misused. Research misconduct can affect the reliability of the scientific literature, and other scientists and journal editors have a clear interest in knowing about it. And the government may be

failing to protect its own interests if the names of parties disciplined by one branch are not available to others.

Making it clear that names will always be kept confidential — as NSERC does — also increases the chance of leaks from understandably frustrated whistleblowers, and of the publishing of names through unofficial routes. Last week, for example, the blog of Retraction Watch,

“Taxpayers have a right to know about instances in which their money has been misused.”

a US watchdog of the scientific literature, guessed at the identity of one of the researchers referred to in the documents that Munro obtained. Such disclosures, although often better than no information at all, can leave the public and other researchers struggling to determine what is true.

NSERC has also redacted the names of institutions, including one that Munro says had to be ‘reminded’ twice to investigate allegations after a researcher had left. Yet there is a clear public interest in knowing whether universities — which typically employ thousands of researchers and educate tens of thousands of students — have received misconduct allegations, and how they have responded.

Redacting the conclusions of an investigation or the actions taken in response to wrong-doing is even worse, arguably serving to protect not only the institutions, but the funding agency itself. For example, one document provided to Munro says that the funding agency has decreed “that [black smudge] be declared ineligible indefinitely to hold or apply for a [black smudge]”. It seems absurd to redact information about how misconduct has been punished. NSERC says that it takes misconduct seriously, that cases in Canada are rare and that its actions send a strong message about the consequences of misconduct. But without any details, the message lacks force. ■

Heritage threats

Research that underpins the preservation of cultural treasures should be championed.

Archaeologists last week announced the discovery of what looks likely to have been a huge shipyard dating back to the second century AD, part of a complex of buildings at Portus, the ancient port of Rome. The discovery (see go.nature.com/qpbirv) is the latest in a major project to explore Portus using an array of geophysical and electromagnetic techniques.

There are any number of similar science-based studies, projects and initiatives under way around the world, all intended to discover or preserve our heritage. As shown in a European web portal launched last week at www.heritageportal.eu, the sciences involved are drawn from across the disciplines. Projects include using climate models to establish which heritage sites are most sensitive to global warming, studies of dust and the effects of its transport in museums and libraries, semantic search techniques to link digital archives and grey literature, and atomic force microscopy to look for signs of decay in paper.

No one could deny that such science is important, given the intrinsic interest of cultural heritage to citizens, not to mention its often sizeable value to economies. But the field of ‘heritage science’ is nevertheless beset by obstacles to its recognition. Because it lacks a disciplinary focus, no traditional university department or disciplinary funding agency readily sees it as their own. Projects are often small scale, and unlikely to reveal the fundamental scientific insights or spin-off products required in competitive funding proposals.

Furthermore, even though national heritage sites and collections may flourish, the science that helps them do so only occasionally

gets the public attention it deserves at the institution that houses it, and rarely, if ever, at the national level. And yet without such science, the value of those sites and collections would be all the more threatened.

To its credit, the European Commission recognizes such problems, and has over the past three years funded the Net-Heritage programme, gathering an overview of national programmes and identifying key weaknesses and strengths. The programme, which finishes at the end of this month, has opened up valuable conversations and coordination between Europe's ministries of culture and other research funders.

What next at the pan-European level? Although the European Commission will continue to have an operational role, the responsibility for coordinated research funding passes to national governments, through a European Joint Programming Initiative (JPI) entitled Cultural Heritage and Global Change. As a result of the Net-Heritage discussions and JPI programmes, there are likely to be collaborations between two or more countries and joint calls for research, in which money does not cross national borders but synergy between parallel national interests is acquired.

However, there are concerns about continuity. The JPI primarily involves collaborations between ministries of research rather than culture, and within each country the culture and research ministries tend not to connect effectively, whether from conflicting pressures or turf protectionism. Accordingly, it is essential that those involved in the JPI, not least the Italian government spearheading it, do whatever they can to prevent such obstacles from undermining achievements in focusing European efforts. Above all, this research will depend on champions at government and institutional levels, as each country

battles with budgetary pressures. Otherwise, the spotlight thrown on heritage science by the Net-Heritage programme and its portal will dim, and the fledgling policy coherence it brings will go to waste. ■

➤ NATURE.COM
To comment online,
click on Editorials at:
go.nature.com/xhunq

CORRECTION

The Editorial 'Heritage threats' (*Nature* **477**, 510; 2011) originally stated that Ireland was leading the European Joint Programming Initiative. In fact, Italy has overall responsibility for coordination. The text has been changed to reflect this.



Why animal research needs to improve

Many of the studies that use animals to model human diseases are too small and too prone to bias to be trusted, says **Malcolm Macleod**.

This is the golden age of medical research. Around the world, scientists are spending more money, writing more papers and building more shiny institutes. Almost all grant applications suggest that a positive funding decision will support research that could lead to new treatments for condition X — usually a growing scourge of modern society.

Many medical discoveries have made real differences to the lives of a great number of people, but could the research be done better?

It seems self-evident that we should encourage high-quality work, but what makes for high quality is a matter of opinion, which hardens over the years into dogma on the assumption that the most established and most venerated got there for a reason, so if one wishes their good opinion then one should do as they did.

Take experiments that use animals to model human diseases. Empirical study of the quality of these experiments is an emerging field, but it does suggest that all is not well. The most reliable animal studies are those that: use randomization to eliminate systematic differences between treatment groups; induce the condition under investigation without knowledge of whether or not the animal will get the drug of interest; and assess the outcome in a blinded fashion. Studies that do not report these measures are much more likely to overstate the efficacy of interventions.

Unfortunately, at best one in three publications follows these basic protections against bias¹. This suggests that authors, reviewers and editors accord them little importance.

Other basic aspects of the design of experiments in animals also receive scant attention. In the face of pressures to reduce the number of animals used, investigators often do studies that are too small to detect a significant effect. To guard against such 'underpowered' studies, researchers should calculate the number of animals required to have a reasonable chance of detecting the anticipated effect given the expected variance of the data. Fewer than one in one hundred such publications report sample-size calculations².

Fewer still define beforehand the most important ('primary') outcome. As a result, they tend to report only the outcomes that happen to show statistical significance, reducing a rigorous, hypothesis-testing experiment to something more like observational research.

The tendency to publish only positive results is another flaw in animal research. Such bias not only prevents scientists from getting credit for high-quality research that happens to be neutral, but also gives a false impression of efficacy. My research has shown that in animal tests of treatments for focal cerebral ischaemia (a model for stroke), publication bias leads to an

overestimation of drug efficacy by about one-third³, increasing risk for both clinical-trial participants and the pharmaceutical industry.

Experimental approaches are not very different throughout the life sciences, so the biases are probably similar too. A scientist's environment is full of potential hazards, such as non-renewal of funding, and potential rewards — getting published and receiving grants. As long as cheap, underpowered studies are more likely to have exciting positive (if false) results than expensive, well conducted, large studies — and as long as journals don't seem to know the difference — the pressure will remain to do what everyone else does.

So we need to change the rules. If publication in high-impact journals continues to be a yardstick, then the review process must do

much more to assess bias. The ARRIVE (Animal Research: Reporting *In Vivo* Experiments) guidelines⁴, endorsed by, among others, Nature Publishing Group, are a good start. But, as Don Quixote observed, the proof of the pudding will be in the eating.

There must also be better ways to publish neutral studies. If the focal cerebral ischaemia literature reflects the life sciences generally, then 16% of studies go unpublished, and tackling publication bias would increase the number of manuscripts published every year by 160,000. At current growth rates we would expect this increase anyway over the next four years, so sorting out publication bias should be possible.

At the very least, we should look for ways to register all experiments — so that investigators can receive credit for work done and so that those seeking to summarize what is known

have access to all relevant data. Such a system could be flexible, with information embargoed for a time to protect intellectual property.

It is hugely distressing to hear highly motivated young scientists say that they would prefer to do their research 'properly', but that if they don't get more published from their PhD work they will never find a postdoc position. They feel forced to lower their standards. We owe it to them to create an environment in which the rewards for conducting high-quality research are more immediately apparent. ■

Malcolm Macleod is a clinical neuroscientist at the University of Edinburgh, UK, and a member of CAMARADES (Collaborative Approach to Meta Analysis and Review of Animal Data from Experimental Studies). e-mail: malcolm.macleod@ed.ac.uk

NATURE.COM
Discuss this article
online at:
go.nature.com/zdqwqt

1. Kilkenney, C. et al. *PLoS ONE* **4**, e7824 (2009).
2. Sena, E., van der Worp, H. B., Howells, D. & Macleod, M. *Trends Neurosci.* **30**, 433–439 (2007).
3. Sena, E. S. et al. *PLoS Biol.* **8**, e1000344 (2010).
4. Kilkenney, C. et al. *PLoS Biol.* **8**, e1000412 (2010).

SEVEN DAYS

The news in brief

RESEARCH

Faster than light?

Physicists were this week rushing to pore over the details of an Italian experiment suggesting that neutrinos can travel faster than light. The experiment, OPERA, is a neutrino detector under the Gran Sasso National Laboratory near L'Aquila. It saw a beam of neutrinos travelling 730 kilometres from CERN, Europe's particle-physics laboratory in Geneva, Switzerland, arrive 60 nanoseconds earlier than light would if travelling in a vacuum. The results were unveiled on 22 September. See page 520 for more.

RIP XMRV

Researchers have partially retracted a controversial study reporting the discovery of a virus called XMRV in blood from patients with chronic fatigue syndrome (V. C. Lombardi *et al.* *Science* **326**, 585–589; 2009). The findings came under attack when other laboratories were unable to reproduce them (see *Nature* **471**, 282–285; 2011). On 22 September, *Science* published the retraction together with the findings of a working group that had distributed patient samples to nine independent laboratories. Only two could find XMRV or related viruses, and even those labs did not reproduce their findings when given replicate samples. See go.nature.com/olxles for more.

POLICY

Nuclear reaction

The International Atomic Energy Agency (IAEA) will be obliged to speak out in times of crisis, under an action plan approved by member states on 22 September. The agency was widely criticized

for failing to give clear and independent assessments of the nuclear accident at the Fukushima Daiichi power plant in Japan in March. The new plan calls on the IAEA to provide “timely, clear, factually correct, objective and easily understandable information during a nuclear emergency”. The agency will also consider reforming its seven-point nuclear-emergency scale. See go.nature.com/c7ecgr for more.

Disease summit

Without committing themselves to specific targets, world leaders pledged action on non-communicable disease at a United Nations summit in New York last week. The meeting emphasized the toll of conditions such as heart

attack, cancer and diabetes (see *Nature* **477**, 260–261; 2011), and discussed actions such as controls on salt and tobacco consumption. It will be up to individual nations to put such controls into place.

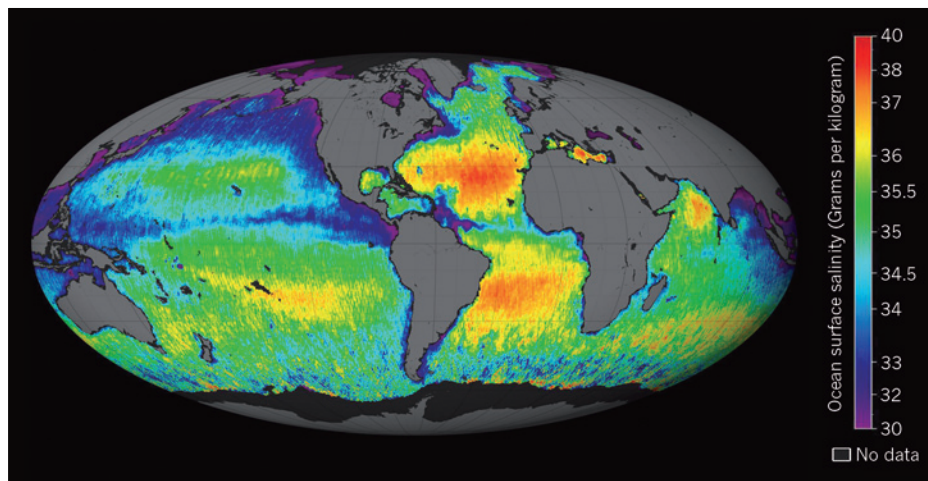
UK science cuts

The Royal Society in London has written to Britain's physical sciences funding agency, telling it to “pause” controversial changes to its funding strategy. The letter, released on 22 September and co-signed by other UK learned societies, came after months of discontent from scientists at how the Engineering and Physical Sciences Research Council was passing on government budget cuts. Researchers complain that the agency has not consulted

widely enough, and that it has not used rigorous evidence to decide on measures such as reducing support for synthetic organic chemistry. The agency's council meets in October to consider the criticisms.

Working balance

The US National Science Foundation, one of the country's main sources of research funding, has announced a plan to encourage more flexible workplace policies. The changes include allowing scientists to delay or suspend grants for up to a year to care for new children or fulfil other family obligations, the agency says. It asked universities and research institutes to adopt similar policies.



SOURCE: NASA/GSFC/JPL-CALTECH

Measuring the ocean's saltiness

This map showing variations in saltiness in the world's oceans is the first produced by NASA's Aquarius satellite, which launched in June. The probe is a collaboration with Argentina's space agency, CONAE, and picks up weak microwave radiation emitted naturally by the ocean. The radiation varies according to the electrical conductivity of the water, which in turn is tied to salinity. Because salinity is linked to water

density and evaporation, the data could help to confirm theories about the global water cycle and its response to climate change. Most of the features apparent in the map are well known — such as the greater salinity of subtropical regions that experience more evaporation from sunlight. The results, released on 22 September, will be combined with data from the European Space Agency's Soil Moisture and Ocean Salinity (SMOS) mission.

PEOPLE

THE HUNTSVILLE TIMES/LANDOV



Murder plea

Amy Bishop, the biologist who shot and killed three colleagues, and injured three others, at the University of Alabama in Huntsville last year (see *Nature* 465, 150–155; 2010) entered a plea of not guilty “by reason of mental disease or defect”, in an Alabama courtroom on 22 September (pictured). Bishop’s trial will begin on 19 March 2012.

Federal fraud

The head of a US biotechnology company that received federal funds to research vaccines has been indicted for fraud. Jian-Yun Dong, president of GenPhar in Mount Pleasant, South Carolina, and an immunologist at the Medical University of South Carolina in Charleston, is charged (along with an unnamed vice-president)

with making “false, fictitious and fraudulent” statements in 2004–11 in order to get research grants from the National Institutes of Health, among other agencies. The indictment — filed at the US district court in Charleston in April and publicly released on 19 September — says that Dong obtained at least US\$3.6 million as a result of his offenses.

Green leader dies

Inspirational environmental campaigner Wangari Maathai died in Nairobi on 25 September, at the age of 71. Maathai was the founder of the African tree-planting initiative the Green Belt Movement, which aims to empower women and improve the quality of their lives by providing better access to clean water and firewood while conserving the environment. A veterinary scientist, Maathai won the Nobel Peace Prize in 2004 for her work in supporting democracy, human rights and the environment.

BUSINESS

Sequencing sags

DNA-sequencing firm Pacific Biosciences is laying off 130 employees, or 28% of its staff, the company said on 20 September. Based in Menlo Park, California, the

firm started to ship its first commercial machines for real-time, single-molecule DNA sequencing in April. But demand has not climbed as rapidly as hoped. Some other sequencing companies, such as Life Technologies in Carlsbad, California, also announced lay-offs this year, as the sequencing market reaches saturation ahead of what analysts expect to be a few difficult years for the budgets of research funders such as the US National Institutes of Health.

Shadows on solar

Republican politicians have ramped up criticism of the US government’s clean-energy loan-guarantee programme, which has come under fire since the bankruptcy of solar-energy company Solyndra, based in Fremont, California, early this month. A House of Representatives subcommittee is asking whether the Department of Energy was too hasty in granting the firm a US\$535-million loan guarantee in March 2009. On 23 September, two Solyndra executives refused to give testimonies at the subcommittee hearing. Meanwhile, US solar firms First Solar, based in Tempe, Arizona, and SolarCity, of San Mateo, California, have blamed red tape for their failures to secure

COMING UP

30 SEPT–2 OCT

The fourth annual meeting on personal genomes, at Cold Spring Harbor, New York, looks at the issue of incorporating genome data into routine clinical practice.

go.nature.com/hozwbo

3–5 OCTOBER

The winners of the 2011 Nobel prizes for physiology or medicine, physics and chemistry are announced in Stockholm.

www.nobelprize.org

3–5 OCTOBER

The world stem-cell summit in Pasadena, California, focuses on progress in translating stem-cell research for the clinic.

go.nature.com/kov6xu

loan guarantees before the end of this fiscal year on 30 September.

Stem-cell trials

The biotechnology firm Advanced Cell Technology has been given the go ahead to conduct a trial that makes use of human embryonic stem (ES) cells in the United Kingdom — the first such approval outside the United States. The company, which is headquartered in Santa Monica, California, is already conducting US trials; the UK trial will be a repeat version, using retinal cells derived from human ES cells to treat people with Stargardt’s disease, an inherited form of progressive blindness. The company says it is in late-stage talks with regulators and clinicians in France, China and elsewhere to launch further global trials. See go.nature.com/ogdjgt for more.

➔ NATURE.COM

For daily news updates see:

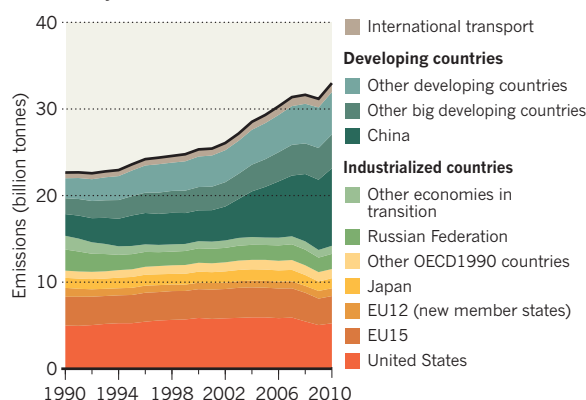
www.nature.com/news

TREND WATCH

Global greenhouse-gas emissions increased by 5.8% in 2010 to hit an all-time high of 33 billion tonnes, as continued growth in developing countries swamped both greater use of renewable power and gains in energy efficiency, according to an analysis by the European Commission’s Joint Research Centre and the PBL Netherlands Environmental Assessment Agency. Emissions in China and India increased by 10% and 9%, respectively, compared with 3% in the United States.

LONG-TERM TRENDS IN CO₂ EMISSIONS

Growth in developing nations is swamping any cuts in greenhouse-gas emissions by industrialized countries.



SOURCE: JRC/PBL

NEWS IN FOCUS

PUBLIC HEALTH Escalating the fight against childhood diarrhoea **p.519**

PHYSICS Improbable neutrino result prompts hunt for cause **p.520**

ETHICS Aboriginal genome addresses consent issues **p.522**



BIOMEDICINE Four fixes for costly, failure-prone clinical trials **p.526**

A. GUPTA/REUTERS/CORBIS



Indian carbon-credit claims — such as those for the Baglihar Dam (above) — are under scrutiny.

CLIMATE CHANGE

Clean-energy credits tarnished

WikiLeaks reveals that most Indian claims are ineligible.

BY QUIRIN SCHIERMEIER

As the world gears up for the next round of United Nations climate-change negotiations in Durban, South Africa, in November, evidence has emerged that a cornerstone of the existing global climate agreement, the international greenhouse-gas emissions-trading system, is seriously flawed.

Critics have long questioned the usefulness of the Clean Development Mechanism (CDM), which was established under the Kyoto Protocol. It allows rich countries to offset some of their carbon emissions by investing in climate-friendly projects, such as hydroelectric power and wind farms, in developing countries. Verified projects earn certified emission reductions

(CERs) — carbon credits that can be bought and sold, and count towards meeting rich nations' carbon-reduction targets.

But a diplomatic cable published last month by the WikiLeaks website reveals that most of the CDM projects in India should not have been certified because they did not reduce emissions beyond those that would have been achieved without foreign investment. Indian officials have apparently known about the problem for at least two years.

"What has leaked just confirms our view that in its present form the CDM is basically a farce," says Eva Filzmoser, programme director of CDM Watch, a Brussels-based watchdog organization. The revelations imply that millions of tonnes of claimed reductions in

greenhouse-gas emissions are mere phantoms, she says, and potentially cast doubt over the principle of carbon trading. "In the face of these comments it is no wonder that the United States has backed away from emission trading," Filzmoser says.

The cable, written on 16 July 2008, was sent by the US consulate in Mumbai, India, to the US secretary of state, and summarizes a discussion of the CDM involving representatives of the consulate and the US Government Accountability Office, along with Indian officials and executives of large Indian companies. At the time, 346 Indian projects had been registered with the CDM's executive board. Today, more than 720 Indian projects have been approved and have gained some 120 million tonnes' worth of carbon credits, a large fraction of the 750 million tonnes issued since 2005 (see 'Cleaning up').

Yet on the evidence of discussions at the meeting, most of the carbon-offset projects in India fail to meet the CDM requirements set by the UN Framework Convention on Climate Change. The cable also describes the UN's validation and registration process as "arbitrary".

Indian authorities were also criticized in the cable. All CDM projects must be validated nationally, then verified independently by an accredited firm. But the cable quotes R. K. Sethi, then chairman of the CDM's executive board and member-secretary of the Indian CDM authority in New Delhi, as admitting that the authority simply "takes the project developer at his word for clearing the additionality barrier".

"This will not invalidate carbon trading, but it does go to show that the CDM has serious flaws," says Mark Maslin, a climatologist at University College London. "In India and China, the multiple levels of governance which you need to have in place to make carbon trading work are simply not there."

Martin Hession, head of global carbon markets at the UK Department of Energy and Climate Change, and chairman of the CDM executive board, says that the critical remarks in the cable date from a time when "people were complaining a lot" about problems with the CDM. Controversies over whether specific projects reduce net emissions are still common, he says. But since 2008, the board has followed more stringent guidelines for verifying the eligibility of projects and for enhancing the overall efficiency of the scheme.

"The CDM is much more transparent and predictable than the tenor of these remarks ►

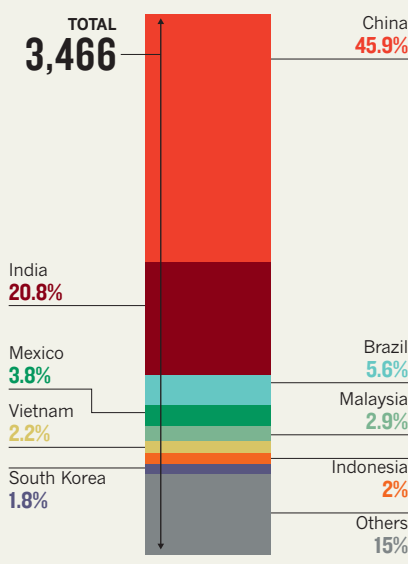
► might suggest," he says. "We reject many projects in India and China because they fail to meet the required criteria, and we do in fact often get the message that project validation has become too stringent."

Others argue that the rules are still not rigorous enough. In some circumstances, the CDM may actually have encouraged the production in developing countries of the coolant chemical HFC-23, an extremely potent greenhouse gas (M. Wara *Nature* **445**, 595–596; 2007). Critics have demanded harsher sanctions against validating companies found guilty of lax oversight, together with clearer conflict-of-interest policies and tighter rules on what qualifies as an additional clean-development project.

International Rivers, an environmental campaigning group based in Berkeley, California, is now calling on the CDM executive board to reject the 412-megawatt Rampur hydro-power project in Himachal Pradesh, India, which is awaiting CDM approval. The project could earn some 15 million carbon credits from 2012 to 2022, amounting to an estimated US\$150-million windfall for the Shimla-based

CLEANING UP

Roughly one-fifth of projects in the Clean Development Mechanism are registered in India — but most do not comply with the scheme.



developer Satluj Jal Vidyut Nigam Limited, the group says. But the decision to finance that project was taken long before the CDM was even created, says Himanshu Thakkar, director of the Delhi-based South Asia Network on Dams, Rivers and People, clearly invalidating its application.

The company stands by its claim that the project qualifies for the CDM, and says that the Indian government approved the investment proposal for the project in 2007, when the CDM was already in place. As *Nature* went to press, the CDM's executive board, which met in Quito, Ecuador, this week, had not yet decided whether to approve the Rampur project.

Despite the controversy, the European Union seems determined to continue its mandatory emissions-trading system, which it sees as crucial in tackling climate change. There's little doubt about the urgency of that goal: global carbon dioxide emissions have increased by 45% since 1990, reaching an all-time high of 33 billion tonnes in 2010, according to a report released last week by the European Commission (see page 515). ■

GENOMICS

Europe to map the human epigenome

DNA-modification studies get a multi-million euro boost.

BY ALISON ABBOTT

The health-research division of the European Commission launches its largest-ever project next week with a €30-million (US\$41-million) investment in understanding the human epigenome, the constellation of DNA modifications that shape how genes are expressed.

With the project, called BLUEPRINT, Europe intends to become a major player in the International Human Epigenome Consortium (IHEC), set up last year to help biologists understand how the epigenome influences health and disease.

All the cells that make up an individual originate from the same fertilized egg and share the same genome. But during development, cells acquire epigenetic changes — such as chemical modifications to DNA, and changes in the shape of its tightly coiled three-dimensional structure — that affect which genes will be active in which cells at a given time and which will be silent.

The epigenome then remains largely stable, so that a liver, for example, remains a

liver throughout life. But small parts of the epigenome change constantly as the cell responds to environmental changes. The epigenome also changes in diseases, including cancer.

The importance of the epigenome in health and disease is becoming increasingly clear. But researchers studying it have faced a big hurdle: the lack of a reliable library of high-quality, quantitative reference epigenomes against which new data can be compared. Little is known, even, about how much the epigenome normally varies between individuals, or between the different cells in an individual.

Enter BLUEPRINT, which unites 41 institutions and more than 50 principal investigators across Europe who will contribute a further €10 million to the project. BLUEPRINT will provide at least 100 reference epigenomes toward the IHEC's goal of amassing 1,000 reference epigenomes by 2020.

BLUEPRINT has chosen to focus on the blood system. This should help move discoveries quickly into the clinic, as many diagnostic tests rely on blood samples. "Blood is also what gets stored in biobanks and used for genome analysis," says BLUEPRINT coordinator

Henk Stunnenberg at the Nijmegen Center for Molecular Life Sciences in the Netherlands.

Beyond these conveniences, there was also a compelling biological rationale for the choice, he says. Unlike most tissues, blood cells are constantly renewed, so blood comprises a mix of cells at different stages of maturity. "Blood-cell epigenomes may reveal some general rules about how cells develop," Stunnenberg says.

BLUEPRINT will generate reference epigenomes from 60 different cell types, taken from the blood of healthy individuals stored in the UK's national blood bank. Each epigenome will include a full genome sequence, and genome-wide quantitative data on the occurrence and distribution of nine different epigenetic markers.

For comparison with healthy epigenomes, the consortium will produce reference epigenomes for more than 60 blood-cancer cell types. It will carry out experiments in mice to work out how much of the epigenome is heritable. The consortium also plans to generate lower-resolution epigenomes from two blood-cell types from 100 healthy people to provide a first quantitative indication of natural individual variation.

"BLUEPRINT is the first big epigenome project to be specifically created in alignment with the IHEC mission," says Peter Jones at the University of Southern California in Los Angeles, who helped to launch the IHEC. "Blood epigenomes are particularly exciting because we know an awful lot about the biology of how blood stem cells differentiate, but little about the sequence of epigenomic events involved in the processes that are going to be relevant for disease." ■

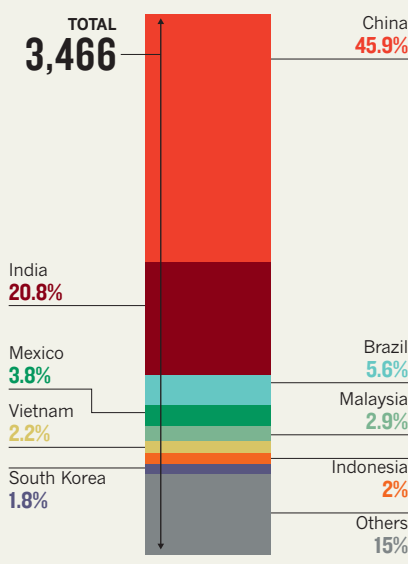
► might suggest," he says. "We reject many projects in India and China because they fail to meet the required criteria, and we do in fact often get the message that project validation has become too stringent."

Others argue that the rules are still not rigorous enough. In some circumstances, the CDM may actually have encouraged the production in developing countries of the coolant chemical HFC-23, an extremely potent greenhouse gas (M. Wara *Nature* **445**, 595–596; 2007). Critics have demanded harsher sanctions against validating companies found guilty of lax oversight, together with clearer conflict-of-interest policies and tighter rules on what qualifies as an additional clean-development project.

International Rivers, an environmental campaigning group based in Berkeley, California, is now calling on the CDM executive board to reject the 412-megawatt Rampur hydro-power project in Himachal Pradesh, India, which is awaiting CDM approval. The project could earn some 15 million carbon credits from 2012 to 2022, amounting to an estimated US\$150-million windfall for the Shimla-based

CLEANING UP

Roughly one-fifth of projects in the Clean Development Mechanism are registered in India — but most do not comply with the scheme.



developer Satluj Jal Vidyut Nigam Limited, the group says. But the decision to finance that project was taken long before the CDM was even created, says Himanshu Thakkar, director of the Delhi-based South Asia Network on Dams, Rivers and People, clearly invalidating its application.

The company stands by its claim that the project qualifies for the CDM, and says that the Indian government approved the investment proposal for the project in 2007, when the CDM was already in place. As *Nature* went to press, the CDM's executive board, which met in Quito, Ecuador, this week, had not yet decided whether to approve the Rampur project.

Despite the controversy, the European Union seems determined to continue its mandatory emissions-trading system, which it sees as crucial in tackling climate change. There's little doubt about the urgency of that goal: global carbon dioxide emissions have increased by 45% since 1990, reaching an all-time high of 33 billion tonnes in 2010, according to a report released last week by the European Commission (see page 515). ■

GENOMICS

Europe to map the human epigenome

DNA-modification studies get a multi-million euro boost.

BY ALISON ABBOTT

The health-research division of the European Commission launches its largest-ever project next week with a €30-million (US\$41-million) investment in understanding the human epigenome, the constellation of DNA modifications that shape how genes are expressed.

With the project, called BLUEPRINT, Europe intends to become a major player in the International Human Epigenome Consortium (IHEC), set up last year to help biologists understand how the epigenome influences health and disease.

All the cells that make up an individual originate from the same fertilized egg and share the same genome. But during development, cells acquire epigenetic changes — such as chemical modifications to DNA, and changes in the shape of its tightly coiled three-dimensional structure — that affect which genes will be active in which cells at a given time and which will be silent.

The epigenome then remains largely stable, so that a liver, for example, remains a

liver throughout life. But small parts of the epigenome change constantly as the cell responds to environmental changes. The epigenome also changes in diseases, including cancer.

The importance of the epigenome in health and disease is becoming increasingly clear. But researchers studying it have faced a big hurdle: the lack of a reliable library of high-quality, quantitative reference epigenomes against which new data can be compared. Little is known, even, about how much the epigenome normally varies between individuals, or between the different cells in an individual.

Enter BLUEPRINT, which unites 41 institutions and more than 50 principal investigators across Europe who will contribute a further €10 million to the project. BLUEPRINT will provide at least 100 reference epigenomes toward the IHEC's goal of amassing 1,000 reference epigenomes by 2020.

BLUEPRINT has chosen to focus on the blood system. This should help move discoveries quickly into the clinic, as many diagnostic tests rely on blood samples. "Blood is also what gets stored in biobanks and used for genome analysis," says BLUEPRINT coordinator

Henk Stunnenberg at the Nijmegen Center for Molecular Life Sciences in the Netherlands.

Beyond these conveniences, there was also a compelling biological rationale for the choice, he says. Unlike most tissues, blood cells are constantly renewed, so blood comprises a mix of cells at different stages of maturity. "Blood-cell epigenomes may reveal some general rules about how cells develop," Stunnenberg says.

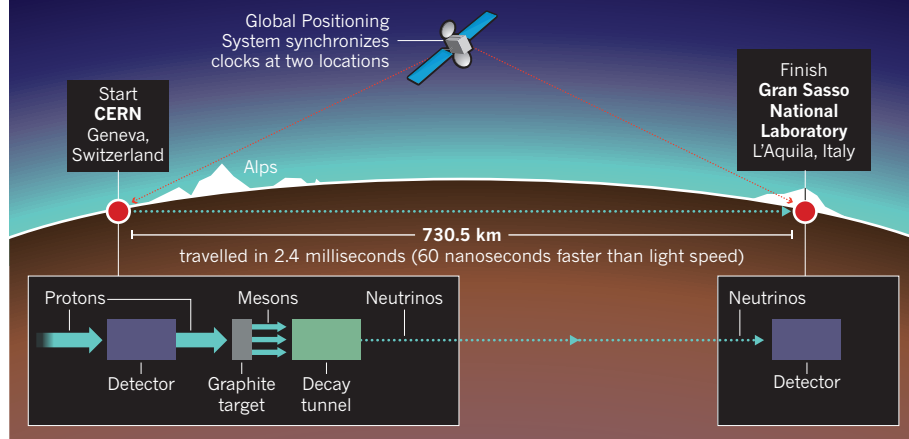
BLUEPRINT will generate reference epigenomes from 60 different cell types, taken from the blood of healthy individuals stored in the UK's national blood bank. Each epigenome will include a full genome sequence, and genome-wide quantitative data on the occurrence and distribution of nine different epigenetic markers.

For comparison with healthy epigenomes, the consortium will produce reference epigenomes for more than 60 blood-cancer cell types. It will carry out experiments in mice to work out how much of the epigenome is heritable. The consortium also plans to generate lower-resolution epigenomes from two blood-cell types from 100 healthy people to provide a first quantitative indication of natural individual variation.

"BLUEPRINT is the first big epigenome project to be specifically created in alignment with the IHEC mission," says Peter Jones at the University of Southern California in Los Angeles, who helped to launch the IHEC. "Blood epigenomes are particularly exciting because we know an awful lot about the biology of how blood stem cells differentiate, but little about the sequence of epigenomic events involved in the processes that are going to be relevant for disease." ■

RACING LIGHT

By comparing the proton signal at CERN to the resulting neutrino signal at Gran Sasso, the OPERA experiment was able to calculate the neutrinos' time of flight as they passed through Earth.



neutrinos to 'oscillate' from one type to others as they travel. It was later adapted to measure neutrino velocity to a higher precision than any previous experiment. A crucial aspect of this was the pioneering use of the Global Positioning System (GPS) to establish a time reference at both ends of the neutrino's flight path. Another was the careful timing of particles through different stages of the experimental apparatus, using a pair of ultra-high-accuracy caesium clocks to determine sources of delay.

In March 2011, the group was shocked to discover that its data suggested neutrinos were arriving 60 nanoseconds faster than light would do over the equivalent distance. "After six months of cross-checking we decided to go public," says Dario Autiero of the Institute of Nuclear Physics in Lyons (IPNL), France, who is the physics coordinator for OPERA. Most members of the collaboration agreed with the decision, but physicist Caren Hagner of the German Electron Synchrotron (DESY) in Hamburg was among those who declined to put their name to the result. In her case, she says, it was not because of any specific problem, but because she would have liked to spend more time checking the finding.

The release of the team's data on the arXiv server (The OPERA Collaboration, preprint at <http://arxiv.org/abs/1109.4897>; 2011) and in a presentation (see go.nature.com/kl4jah) now have the neutrino community looking for sources of error that might explain the result. Two elements of the experiment receiving particular scrutiny include the GPS-synchronization system and the profile of the proton beam that generates the neutrinos as a by-product of colliding with a target. Experimenters determined the flight time by comparing the shape of the proton signal at CERN to that of the neutrino signal received at Gran Sasso (see 'Racing light').

Two other collaborations are looking to check the OPERA result independently. Neither can yet time neutrinos to the same level of precision; however, an upgrade is already planned for MINOS, which sends a beam of muon neutrinos from Fermilab to the Soudan mine in Minnesota — roughly the same distance as from CERN to Gran Sasso.

Japan's T2K experiment, which sends neutrinos 295 kilometres from an accelerator in Tokai to the Super-Kamiokande detector in Kamioka, was shut down after the earthquake on 11 March. Scientists there are also discussing an upgrade to check OPERA's result, says co-spokesman Chang Kee Jung.

The upgrades would take more than a year to implement. In the meantime, researchers with both the T2K and MINOS collaborations are taking a second look at their existing data to see whether they are consistent with OPERA's result. Plunkett says that the MINOS group might have an answer within a few months. And even if the speed of light remains unbroken, a move to more accurate timing will bolster experiments in the long run. ■

PARTICLE PHYSICS

Speedy neutrinos challenge physicists

Experiment under scrutiny as teams prepare to test claim that particles can beat light speed.

BY EUGENIE SAMUEL REICH

The joke begins with the barman saying: "I'm sorry, we don't serve neutrinos." Then the punch line: a neutrino walks into a bar.

Such causality-bending humour has been rife on the Internet in the past week, following the news that an experiment at the Gran Sasso National Laboratory near L'Aquila, Italy, has apparently clocked neutrinos exceeding the speed of light as they travelled 730 kilometres from their source at CERN, Europe's particle-physics laboratory near Geneva, Switzerland.

The finding by the OPERA (Oscillation Project with Emulsion-tracking Apparatus) collaboration, released on 22 September, has the media abuzz with talk of a century's

worth of physics upended, starting with Albert Einstein's special theory of relativity. This sets the velocity of light as the inviolable and unattainable limit for matter in motion, and links it to deeper aspects of reality, such as causality.

Physicists, for the most part, suspect that an unknown systematic error lies behind OPERA's startling result. But nothing obvious has emerged, and many see the experiment as a tour de force because of its high precision. "It is quite a complicated experiment but they did a professional job," says Rob Plunkett, co-spokesman for the MINOS (Main Injector Neutrino Oscillation Search) experiment at Fermilab in Batavia, Illinois, which is likely to investigate the claim.

OPERA was switched on in 2006 to study the peculiar ability of the fleeting, nearly massless



TOP STORY



Bat virus poses epidemic threat to people in West Africa
go.nature.com/usp7cw

MORE NEWS

- Texas holds firm on physics cutback go.nature.com/t1nigm
- Ecologists dig deep for DNA go.nature.com/ya8ien
- Iconic exoplanet may be a mirage go.nature.com/iyjgxi
- Nitrate levels soar off China's coast go.nature.com/zk7nyd

INTELLECTUAL PROPERTY

'Patent trolls' target biotechnology firms

Opportunistic patent litigation threatens the life sciences.

BY ERIKA CHECK HAYDEN

SOURCE: REF. 2

The biotechnology industry has had its share of woes, but so far 'patent trolls' have not numbered among them. These companies, which profit by legally enforcing patents they own rather than developing products, may benefit from a 31 August ruling at a US federal court of appeal in Washington DC.

The court upheld a lawsuit filed by Classen Immunotherapies of Baltimore, Maryland, against four biotechnology companies and a medical group, for infringing on a patent that covered the idea of trying to link infant vaccination with later immune disorders. A district court had thrown out the lawsuit, finding that the concept at the heart of the case amounted to an abstract idea that could not be patented. The appeals court found otherwise.

Beyond its complex particulars, the case sets "a troubling precedent", says James Bessen, a lawyer at the Boston University School of Law, Massachusetts, "because you're patenting something that's very broad". (The patents include the act of reading the published scientific literature and using it to create vaccination schedules that minimize immune disorders.)

Joseph Zito, the plaintiff's lawyer, says that his client, John Barthelow Classen, was the first researcher to connect vaccination schedules to immune disorders through animal studies. "He doesn't want to stop anybody," Zito says. "He wants to make sure people use vaccines safely." But very broad patents have posed a problem in the technology field, where some firms amass vast portfolios of patents bought up from inventors and look for targets to sue.

HARMING SOCIETY

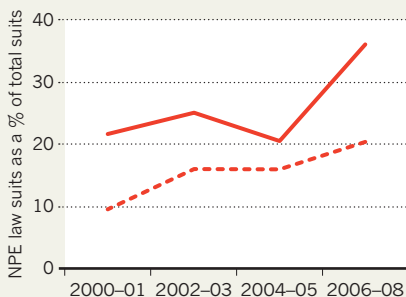
In a study released on 19 September, Bessen and his co-authors estimate that patent lawsuits filed by trolls, also known as 'non-practicing entities' (NPEs), caused defendants to lose more than US\$500 billion in value from 1990 to 2010 (ref. 1). The figure was calculated by analysing the effect on defendants' stock prices when they were sued.

"We conclude that the loss of billions of dollars of wealth associated with these lawsuits harms society," the authors write.

In 2009, Colleen Chien, a lawyer and patent expert at Santa Clara University in California, calculated that the proportion of patent-litigation cases involving NPEs grew

COURTING CHAOS

US patent litigation involving 'non-practicing entities' (NPEs) is rising (dashed line) as is the proportion of defendants in NPE cases (solid line).



from 10% to 20% between 2000 and 2008. The proportion of defendants whose cases involve NPEs — which is disproportionately high — rose from 22% to 36% in the same period² (see 'Courting chaos'). Roughly two-thirds of all suits are over software patents, but experts warn that the life sciences are not immune.

"There is definitely reason for concern in the biotech field," says Julie Samuels, a staff attorney at the non-profit Electronic Frontier Foundation in San Francisco, California.

So far, Chien says, biotechnology companies have not been as vulnerable because their patents are not as broad as software patents, and require more time and expertise to develop. Yet the Classen case shows that broad patent claims can survive in court. And some of the largest patent holders are showing signs that they are interested in the biotechnology sector. For example, Intellectual Ventures of Bellevue, Washington, founded by Nathan Myhrvold, a former Microsoft executive, owns hundreds of patents in the biotechnology industry. The company says it has no plans to litigate in the field.

A weak economy could also become a factor if it leads biotech companies to sell off their patents to opportunistic litigators. And if companies start profiting from troll behaviour, the practice will spread, says lawyer Robert Cook-Deegan of Duke University in Durham, North Carolina. "If people do it and get away with it, it sets up an incentive for other people to behave the same way," he says. ■

1. Bessen, J. E. et al. *Boston Univ. School of Law, Law and Economics Research Paper* 11-45 (2011).
2. Chien, C. V. *North Carolina Law Review* **87**, 1571-1615 (2009).

Aboriginal genome analysis comes to grips with ethics

Sequencing effort provides a model for future studies of museum samples.

BY EWEN CALLAWAY

En route from Sydney to Perth, Australia, in the early 1920s, British ethnologist Alfred Cort Haddon acquired a tuft of human hair from a young Aboriginal man. He added it to his sizeable collection of hair from people living around the world.

Ninety years later, those locks have yielded the first complete genome sequence of an Aboriginal Australian, and provided clues about the timing of human migrations from Africa to Asia¹ (see 'Early human explorers charted a bold course'). The work has also underscored the bioethical dilemmas involved in plumb-ing the genomes of indigenous populations — especially when the DNA comes from an archived specimen such as Haddon's. "To be sequencing DNA from the hair of a deceased indigenous person is uncharted ethical territory," says Emma Kowal, a cultural anthropolo-gist at the University of Melbourne.

The genome project, led by Eske Willerslev at the University of Copenhagen, received approval for the work from a group that represents Aboriginals in the region in which the man probably lived. But some scientists are jittery about how others in the Aboriginal community might receive the project, and worry that it could set back efforts to engage Aboriginals in genetic research. "In a sense, every Aboriginal Australian has had something about themselves revealed to the world without their consent," says Hank Greely, who directs the Center for Law and the Biosciences at Stanford University in California.

Aboriginal Australians endured centuries of repression by European colonists, but their wariness of genetic research owes much to the Human Genome Diversity Project (HGDP). This 1990s international collaboration aimed to catalogue the genetic diversity of populations worldwide, but sparked concerns that indigenous peoples were being subjected to neocolonial 'bioprospecting'. "Probably the strongest opposition we ran into anywhere in the world" was in Australia, says Greely, who was an ethical adviser to the project. Plans to include Aboriginal Australian

"Sequencing DNA from the hair of a deceased indigenous person is uncharted ethical territory."



P. TWEEDIE/PANOS

Today's Australian Aboriginals are descended from the first humans to travel far from their African origins.

DNA were eventually scrapped, and the furore's impact continues to reverberate, says Kowal. "The damage that the HGDP has done for the prospect of doing genetic research with Aboriginal people has been significant." Researchers who work with Aboriginal Australians are now expected to obtain consent not only from the individuals concerned, but also from local and sometimes state-wide groups representing Aboriginal communities across Australia.

Sheila van Holst Pellekaan, a geneticist at the University of New South Wales in Sydney, says that she has encountered long delays in getting continued approval for her genetic studies of ancestry and disease among Aboriginal Australians, and can share only limited data with other researchers. Such obstacles, she says, mean that Aboriginals, who suffer from high rates of diabetes and heart disease, may miss out on advances in genetic medicine.

Willerslev says that he was not fully aware of these issues when his team set out to sequence Haddon's sample. Moreover, the sample came with little identifying information, so Willerslev's team had no idea whom to ask for permission to study it.

A Danish bioethical review board did not believe it was necessary to review the project because it viewed the hair as an archaeological specimen and not a biological one, Willerslev

says. However, after his team sequenced the genome, an Australian colleague put Willerslev in touch with the Goldfields Land and Sea Council, a body based in Kalgoorlie, Western Australia, that represents the 5,000 or so Aboriginal Australians living in the region where Haddon collected the hair sample. In June, Willerslev flew to the region to describe his project to the organization's board and to seek its approval. He says that if the board had rejected his proposal, he would have ended the project and left the genome unpublished.

Craig Muller, the research manager at Goldfields Land and Sea Council, says that he researched the history of the man's hair sample and concluded that the man had donated it freely to Haddon. "Once that was made clear, nobody was bothered that there might be some underlying problem," says Muller.

"I hope that our study will set a precedent on how to potentially deal with these issues," says Willerslev, whose team has obtained similar approval to sequence the genome of the iconic Native American Sitting Bull from a lock of hair.

Despite Willerslev's efforts, "I would suggest there would be a certain amount of unrest in the indigenous communities," says van Holst Pellekaan. Greely agrees that Willerslev's team should have reached out to other Aboriginal groups.

"I think they did everything anyone could reasonably expect them to," counters Mark Stoneking, a molecular anthropologist at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany. He published a complementary analysis of Aboriginal genomes last week², using DNA samples obtained by other scientists with the consent of the Aboriginal Australian individuals involved.

The study also raises broader consent issues over body parts of indigenous people held in museums, says Kowal. Many collections are returning bones to these groups, but the British Museum in London, for instance, generally excludes hair and nails from its repatriation policy. Such specimens are a valuable tool for studying the genomes of people from around the world, including populations that no longer exist, argues Willerslev.

Having a set of widely accepted guidelines for studying such samples would help to guide researchers, journals and funding agencies, says Stoneking. "Hopefully some sort of standards can be developed so everyone feels comfortable going ahead with this research," he says. ■

1. Rasmussen, M. *et al. Science* <http://dx.doi.org/10.1126/science.1211177> (2011).
2. Reich, D. *et al. Am. J. Hum. Genet.* <http://dx.doi.org/10.1016/j.ajhg.2011.09.005> (2011).
3. Reich, D. *et al. Nature* **468**, 1053–1060 (2010).

WAVES OF MIGRATION

Early human explorers charted a bold course

The first complete genome of an Aboriginal Australian is a portrait of a pioneer. Comparisons with other genomes indicate that Australia's original settlers parted company with other humans more than 60,000 years ago, suggesting that they were among the first to venture far beyond Africa¹.

Like other populations outside Africa, the genome carries small chunks of DNA that originally came from Neanderthals. More surprisingly, it shows that Aboriginals' ancestors also interbred with another archaic human population known as the Denisovans, only recently identified from 30,000–50,000-year-old DNA recovered from a finger bone found in a Siberian cave³. Until now, Papua New Guineans were the only modern humans whose ancestors were known to have interbred with Denisovans.

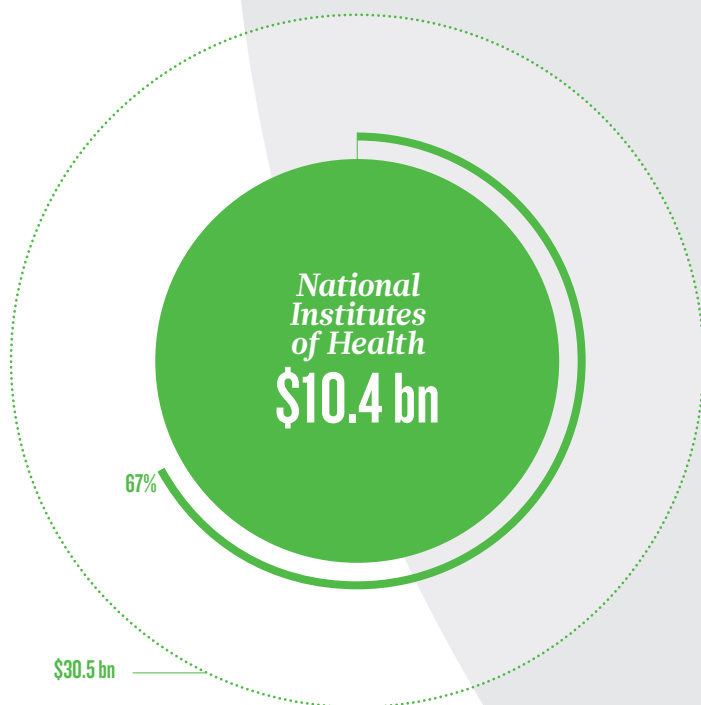
A separate study published last week, based on genetic markers in modern-day Aboriginal Australians, Asians and Pacific Islanders, paints a fuller picture of the Denisovans' genetic legacy. The research, led by Mark Stoneking at the Max Planck

Institute for Evolutionary Anthropology in Leipzig, Germany, revealed a patchwork in which some populations, including Australian Aboriginals, bore varying levels of Denisovan DNA, whereas many of their neighbours, such as the residents of mainland Southeast Asia, contained none².

Stoneking says that this pattern suggests that an early wave of migration into Asia included the ancestors of contemporary Aboriginal Australians. This was followed by a second wave that gave rise to the current residents of mainland Asia. Some members of the first wave interbred with Denisovans. By the time the second wave of Asian migrants arrived, the Denisovans may have vanished. The study also suggests that the Denisovans' range once extended to Southeast Asia and perhaps Oceania.

"Put together, these two papers make an overwhelming case for multiple waves of migration," says David Reich, a population geneticist at Harvard Medical School in Boston, Massachusetts, and an author on the second study. **E.C.**

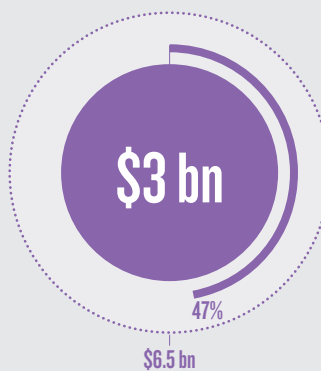
Stimulus funding
across entire
US government
\$840 billion ▶
(85% spent to date)



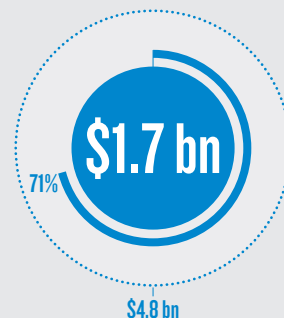
SPENDING A WINDFALL

In 2009, the US government allocated more than US\$52 billion in stimulus funds to research and development. Three agencies received most of the funds for academic research. Much of the money has yet to be spent.

National
Science
Foundation



Office
of Science



Allocated
stimulus
funding

2009 budgets
(not including
stimulus)

Percentage of
stimulus funds
spent to date



STIMULUS—RESPONSE

The United States' 2009 financial stimulus bill has provided research with breathing space, rather than the sharp shot in the arm that many anticipated.

When Cathy Lord arrived in New York last month, her US\$1.2-million grant was about to run out. The autism researcher had won the funding from the National Institutes of Health (NIH) under the American Recovery and Reinvestment Act (ARRA), better known as the stimulus bill, and the money had helped to support her 15-strong team at the University of Michigan in Ann Arbor for two years.

But the expiry of such a major grant didn't leave Lord's team in the lurch — far from it. The grant was not like a standard NIH award. Instead of paying for experiments to test a hypothesis for publication — something that usually takes at least three years — the money allowed Lord's team to compile its existing knowledge, preparing a kit that other researchers can use to screen people for social and behavioural indicators of autism. In that sense, the grant was more of a bonus than a life-line. Now, Lord is about to open an autism clinic and research centre at NewYork-Presbyterian Hospital in White Plains, and about half of her team have followed her there. For them, she says, the end of the ARRA money is no different from the usual rough and tumble that results from the coming and going of research grants. "If you find good people you just try and keep the ball rolling," she says.

Such a smooth transition is not what many expected for scientists facing the end of a stimulus grant. When the bill — now estimated

BY COLIN MACILWAIN

to be worth \$840 billion — was signed into law by President Barack Obama in February 2009, it contained

more than \$52 billion for research and development (see 'Spending a windfall'), of which \$15 billion was expressly for scientific research at the NIH, the National Science Foundation and the Department of Energy's Office of Science (see *Nature* **461**, 856–857; 2009). The stimulus package as a whole was designed to create jobs and ease the pain of the recession, and at first the administration pledged to get this money distributed and spent as quickly as possible — the NIH, for example, devoted much of its allocation to two-year grants that were meant to run from October 2009 to this month. So even as scientists scrambled to get a share of the windfall, some universities and science lobby groups were warning of a 'funding cliff' when the money — which would have boosted total NIH funding by about one-sixth if it was spent over two years — came to an abrupt end.

THE CLIFF THAT WASN'T

From the beginning, however, the funding pulse didn't have quite the anticipated effects. Duke University in Durham, North Carolina, for example, expected a hiring rush after it attracted \$210 million in ARRA funds, making it one of the ten most successful universities in the country in this regard, says Jim Siedow, Duke's vice-provost for research. But

staffing barely budged. “We gave a party that nobody came to,” he says. “A lot of people used the money to keep the people they already had.”

And despite the early political pressure to get the money out of the door quickly, agencies have allowed funds to be released gradually, to avoid waste. The NIH quietly relented earlier this year and is permitting ‘no-cost extensions’ that allow team leaders, including Lord, to defer spending and wind down their projects neatly. The agency has now spent \$7 billion of its \$10.4-billion allocation, and Sally Rockey, the NIH’s deputy director of extramural research, says that “a big chunk of it will still be ongoing” until 2013.

The result, at Duke as elsewhere, has been that the ARRA funding will help to sustain research departments for at least another two years. Siedow says that a detailed internal analysis suggests that Duke will lose only “a very small number” of academic positions as the funding winds down. “We’d been concerned that funding would fall off a cliff; but so far, it hasn’t happened,” he says.

MEASURED RECOVERY

Other research funding agencies report a similarly gradual tail-off. The Department of Energy’s Office of Science says that it expects substantial spending to continue for some time — it expects to disburse about \$260 million in fiscal year 2012 and \$170 million in fiscal year 2013. The NSF, which was under less pressure than the NIH to rush the money out, has so far distributed most of it to previously unsuccessful applicants for standard 3–5-year NSF grants. Only \$1.4 billion of the agency’s ARRA allocation has been spent so far.

Despite this stretching out and morphing, the ARRA is set to leave a distinct legacy in a few areas. One of these is metrics. Kei Koizumi, assistant director for federal research and development at the White House Office of Science and Technology Policy, says that ARRA funding is “the most extensively documented set of research investments we’ve ever seen”.

Because of the initial expectation that stimulus funding would generate jobs, its recipients are required to report regularly on the employment impact of their funding. Their returns are compiled and posted immediately on a public website (www.recovery.gov); these responses suggest that the NIH ARRA funding, for example, is currently supporting the equivalent of about 20,000 full-time positions. However, such self-reported data are not always accurate.

The passage of the stimulus bill also gave birth to Star Metrics, an ambitious cross-government programme set up to monitor research inputs and outputs including, in its first phase, jobs (see *Nature* 465, 682–684; 2010). Star Metrics started off tracking the employment impact of ARRA and non-ARRA research spending at about 70 volunteer research universities, representing an estimated 40% of the US university research system. The results of this analysis, which takes its data directly from university payroll systems, are set to be published shortly.

But further detail about the impact of ARRA research spending may be slow to emerge. Congressional committees have shown little interest so far in raking over the data, according to research lobbyists and university leaders. “Priorities change very rapidly now,” says Chris King, a staff member with the House of Representatives science committee. Formal programme evaluation, as envisaged by Star Metrics, is just “not very critical for people on Capital Hill”, he says.

In the meantime, Star Metrics has shifted its focus away from the ARRA, towards the impacts of all research. “What has happened is that the notion of accountability has spread out much more broadly,” says Julia Lane, an economist and the NSF official who leads Star Metrics. The programme is planning a second stage, in which every researcher in the United States who receives federal funds is assigned an ‘identifier’,

opening the road to a national database that keeps tabs on everything from total federal funding and employment to publications, citations and patent applications.

Although much of the ARRA research money has gone on grants, assumed to be the fastest way to spend the money, some of it is building a concrete legacy in the shape of investment in bricks, mortar and instrumentation. These include a handful of stand-alone projects — such as a \$16-million lab-animal facility at Duke — and many injections into projects already under way.

One of the largest such investments is the \$150 million being spent by the Department of Energy to speed up construction of the \$1-billion National Synchrotron Light Source II: a high-energy photon source at Brookhaven National Laboratory on Long Island, New York, that will enable materials scientists and structural biologists to look at molecules and crystals at the nanoscale. The project was approved in 2005, but the energy department lacked the funds to build it quickly; the facility is now expected to be finished in 2014. “ARRA enabled us to pull forward the construction work, and we’ll have a more fully outfitted facility, with better beamlines and laboratories,” says Steve Dierker, head of photon science at Brookhaven. The facility will be 10,000 times brighter than the light source that it replaces, says Dierker.

The stimulus has also funded experiments in approaches to research funding. Some of these set out specifically to support ‘high-risk’ research projects that might struggle to attract support through the normal grant peer-review process — most notably at the Department of Energy’s new branch, the Advanced Research Projects Agency—Energy (ARPA-E), which was conceived in 2004 but got funded, to the tune of \$400 million, only under ARRA. ARPA-E has been backing research in untested areas, such as the development of microbes engineered to produce biofuels.

More commonly, as at the NIH, ARRA has forced a compression of the peer-review process, and resulted in grants that last for less time than normal. Some of these, such as Lord’s, allow research groups to expand or wrap up projects, but not necessarily test a new hypothesis. “We did an experiment with two-year grants, and we’ll follow up to see how that worked,” says Rockey.

One of the biggest legacies of the science stimulus, however, will be that thousands of students have stayed in science longer than they might otherwise have done.

“WE GAVE A PARTY THAT NOBODY CAME TO.”

“ARRA has provided postdocs and graduate students with great training that’ll help them with the rest of their careers,” says Rockey. “These people will move on to do productive and wonderful things.”

Another legacy will be the ideas, large and small, that have been germinated by ARRA funding — even if their creators are uncertain that they can bring them to fruition. Dan Gauthier, a physicist at Duke, used his three-year, \$480,000 NSF grant to study the interaction of laser light with very cold rubidium atoms. Gauthier is delighted to have had the chance to pursue a project that would otherwise have been out of reach; he thinks that his findings could eventually help in the development of quantum communication networks.

At the same time, he worries about next summer, when the grant supporting him and his two research assistants will end. To keep his team intact, he will have to win normal funding from the NSF. “If we don’t get funded, it’ll just be devastating,” he says. “This was all predicated on the idea that the economy would get better after two or three years: now we don’t know that it will.” ■

Colin Macilwain is a freelance writer based in Edinburgh, UK.

► NATURE.COM

For a breakdown of the 2009 stimulus spending, see: go.nature.com/xlxeel

4

WAYS TO FIX THE CLINICAL TRIAL

Clinical trials are crumbling under modern economic and scientific pressures. Nature looks at ways they might be saved.

BY HEIDI LEDFORD

Developing a drug is a costly gamble. Getting one to market takes, on average, more than ten years and a billion dollars. About 85% of therapies fail in early clinical trials. And of those that survive through to phase III, generally the last step before regulatory approval, only half will actually be approved.

Although a promising compound can fail for many reasons, from safety concerns to corporate decisions, many say that a significant number of good drugs are being lost to outdated and impractical clinical-trial designs (see ‘The clinical-trial cliff’). The drugs may work, says Lillian Siu, an oncologist at Princess Margaret Hospital in Toronto, Canada, “we just don’t know how to test them appropriately.”

Solving the problem may require

fundamental changes to the clinical-trial system to make it faster, cheaper, more adaptable and more in tune with modern molecular medicine. The old paradigm, established in the 1960s, was based on single trials, carried out at single sites, and designed to answer a single question, says Rachel Sherman, associate director for medical policy at the Center for Drug Evaluation and Research at the US Food and Drug Administration (FDA) in Bethesda, Maryland. “But that’s not the world we’re living in now.”

Today, the world is more risk-averse, and demands larger trials to pinpoint safety concerns. Compounds that confer only small benefits when compared with existing drugs require large sample sizes for the results to be statistically significant. As a result, trials have become bigger, and often occur at

multiple sites, even in multiple countries, and can involve thousands of personnel.

The long-heralded era of personalized medicine — tailoring treatments or combinations of treatments to a specific patient — adds its own complications. Such an approach has the potential to lower the failure rates of investigational drugs by testing them only in the individuals most likely to benefit. But research teams struggle to identify biological markers that can be used to stratify patients by the characteristics of their disease, and when they do, diseases can get splintered into rare subtypes that each affect just a few individuals. This means that researchers must screen a much larger pool of potential participants.

Some researchers are working to improve the fortunes of potential therapeutics, however. Here are four ways that they have come up with to give clinical trials a better success rate.



RECRUIT EARLY

Patient recruitment has been a major stumbling block. At least 90% of trials are extended by at least 6 weeks because investigators fail to enrol patients on schedule. Only about one-third of the sites engaged in any multicentre study ever manage to enrol the requisite number, says Kenneth Getz, an expert on clinical research at the Tufts Center for the Study of Drug Development in Boston, Massachusetts. The result: longer, more expensive trials — some of which may never be completed.

Personalized medicine, which has moved apace in cancer research and development, exacerbates the recruitment problem. George Sledge, an oncologist at Indiana University in Indianapolis, offers the example of a class of drug that inhibits enzymes called kinases. Imagine a trial that targets two kinases, one of which is mutated in 25% of patients, and the other of which is mutated in 8%. “You’d have to screen about 50 patients to find one eligible for the trial,” he says. For more complex combinations and more diverse patient pools, the problem “gets to be virtually insoluble”, he says.

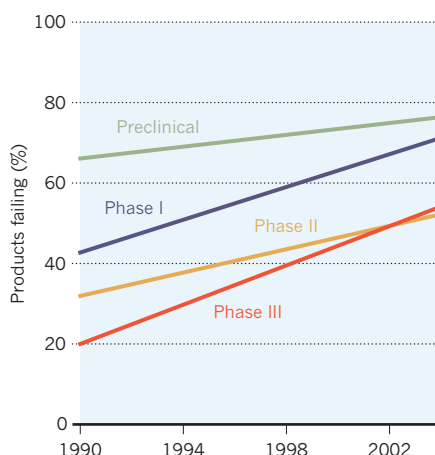
Researchers and patient advocates are trying to make it easier to find eligible volunteers. They are taking a page from organizations such as the Alpha-1 Foundation in Miami, Florida, which has created a registry of patients with alpha-1 antitrypsin deficiency, a disorder that makes them particularly susceptible to lung and liver diseases, who are willing to be contacted about clinical trials.

The Moffitt Cancer Center in Tampa, Florida, for example, runs a Total Cancer Care programme that unites 18 hospitals, compiling medical history, tissue samples and genetic information about each patient’s tumour. Samples are all stored for future analysis, and patients can consent to doctors contacting

THE CLINICAL-TRIAL CLIFF

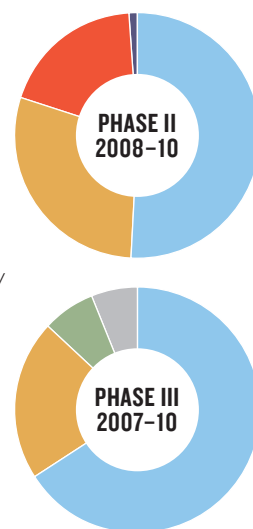
Drug companies are removing more compounds from the pipeline at all levels of testing than ever before.

For projects started between 1990 and 2004, the United States, Europe and Japan have seen sharp rises in the attrition of drugs tested in trials.



Most of the product failures in phase II and III trials are because researchers are unable to demonstrate efficacy or sufficient safety.

■ Efficacy
■ Safety
■ Strategic
■ Pharmacokinetics/bioavailability
■ Commercial/financial
■ Not disclosed



them about trials. A similar union of four institutions is in the works in the Netherlands, says Jan Schellens, an oncologist and pharmacologist at the Netherlands Cancer Institute in Amsterdam.

Negotiating agreements across institutions can be tricky. “Each brings its own values, preferences and interpretation of the privacy laws to the table,” says Walter Kernan, a neurologist at Yale University in New Haven, Connecticut.

“To set up a system in which a hospital provides investigators at another institution with names of patients and diagnoses requires enormous trust.”

Several groups that have been talking about forming networks are likely to face technical barriers. They need to develop appropriate patient-consent forms, unified databases and ultra-secure networks to connect hospitals. “Many of these groups don’t have anywhere near the information-technology networks you’d need to carry this off,” says Sledge. But that doesn’t mean it won’t happen, he adds. “It’s the wave of the future, but it’s not going to be simple.”



SKIP ANIMALS

In 2009, researchers at the National Cancer Institute (NCI) in Bethesda faced a dilemma. They had evidence that blocking a protein called

AKT, involved in cell death and proliferation, could stave off cancer, but initial animal studies on a compound that blocked the protein suggested that the drug was poorly absorbed by the human body. So researchers developed five formulations of the drug. The question was, should they test each one in animals, or could they skip ahead and evaluate them in humans?

In 2006, the FDA and the European Medicines Agency (EMA) introduced guidelines

for testing very small ‘microdoses’ of drugs in humans. These are concentrations less than a one-hundredth of the therapeutic dose. Because the concentrations are so low, the drugs can be tested in a small number of patients without the level of safety data normally required before a phase I study. These early tests, dubbed ‘phase 0’ studies, would show, for instance, how the drug is distributed and broken down in the body, and whether it hits the right molecular target.

Paul Limburg, a gastroenterologist at the Mayo Clinic in Rochester, Minnesota, decided to launch a phase 0 study of the new AKT drug formulations. The results, published in March, allowed the team to pick the one absorbed best by the body for use in future trials (J. M. Reid *et al. Cancer Prev. Res.* 4, 347–353; 2011).

Proponents of phase 0 testing argue that it makes sense to get human data quickly. About one-quarter of the molecules entering clinical trials fail because of “poor pharmacology”, says Schellens. The drug may not be readily absorbed, for example, or may not reach its target organ. With a simple test in humans, he says, “you could kill that drug much earlier, before you have invested so much time and money”.

But Razelle Kurzrock at the MD Anderson Cancer Center in Houston, Texas, is concerned that developers will kill promising drugs in response to negative microdose data, even though the drug may work at higher, therapeutic concentrations. To test this hypothesis, she and her colleagues are conducting phase 0 trials with FDA-approved drugs to find out whether any would have failed at that stage.

And phase 0 trials may be tiny, but they are not easy. They require a test sensitive enough to detect the minute quantities of the drug in the body — and possibly also ways to track its mechanism of action. For veliparib, a ▶

potential anti-cancer drug, Abbott Laboratories, headquartered in Abbott Park, Illinois, started developing assays that were robust and sensitive enough to detect tiny concentrations of the drug more than a year before its researchers embarked on a phase 0 trial, says James Doroshow, the NCI oncologist who led the trial.

Few, says Schellens, plan so far ahead. "Often when people start thinking about the phase 0 trial in the drug development process, they are too late."



USE MODELS

Although phase 0 trials could help to wean researchers off pharmacology studies in animals, moves are also afoot to bring mouse experiments closer to the clinic.

The problem with most of today's animal studies, says Eric Holland, a neurosurgeon at the Memorial Sloan-Kettering Cancer Center in New York, is that they typically focus on safety rather than efficacy. And they rarely predict exactly what will happen in the clinic, because the doses, formulations and schedules of medication differ from those given to the animals. A person with prostate cancer, for example, does not usually take an experimental medication until the standard hormone therapy has failed. But the mice that experimental drug was tested on are unlikely to have received the same treatment.

Pier Paolo Pandolfi, a cancer researcher at Harvard Medical School in Boston, has therefore pioneered a technique called the 'co-clinical trial', in which mice with similar disease characteristics are treated in a similar way to the humans. Holland recently used

be given the same treatment as the patient, so that researchers can track how the tumour cells respond, and use this information to tailor future treatment for the patient.



ALTER COURSE

Perhaps nowhere in science is ignorance prized as highly as in clinical trials. To keep expectations or biases from inadvertently influencing

the results, patients and investigators are often kept in the dark about who is receiving what treatment until the end of the study. At best, an external committee may take a secret peek at the results mid-trial to make sure it is safe for the experiment to continue.

But looking just once at the data is "like driving home from work and only opening your eyes once to see where you're going", says Donald Berry, a statistician at the MD Anderson Cancer Center.

Berry specializes in designing 'adaptive' trials, which can change course as the data roll in. If one treatment regimen seems to be more successful, for example, researchers might increase the proportion of participants that should receive that treatment. These trials can also be used to identify biological markers, such as mutations or altered metabolite levels, associated with the success or failure of a given regimen.

For example, before women enrol in an ongoing adaptive trial called I-SPY2, run by Berry and Laura Esserman, an oncologist at the University of California, San Francisco, their breast-cancer tumours are biopsied and tested for genetic markers. The genetic make-up of their tumours determines whether they are eligible to enter the trial, and which group of

One issue is statistical: the more tests on the data a researcher conducts, the more likely it is that they will introduce false positive results. Scott Evans, a biostatistician at the Harvard School of Public Health in Boston, is more concerned about bias, however. Even if it is an independent panel that views the data and adapts the trial, it is often impossible to hide a change in protocol from everyone. "Any in-flight adjustment you make in a clinical trial is potentially observable," he says. "Now you've created the potential for an operational bias."

"There has been a lot of controversy about adaptive trials," Siu says. "But as the molecular era arrives, they will become more and more relevant."

TEAR DOWN OR TWEAK?

While researchers dream up new, better ways to design clinical trials, many involved acknowledge that the changes with the biggest impact will probably be more bureaucratic than conceptual.

Simply standardizing the forms used to record clinical-trial data would reduce costs and cut down on record-keeping errors and omissions. The NCI is now rolling out a data-management system that will standardize data entry across all 2,000 sites that conduct NCI-sponsored trials, says Doroshow. The FDA is also looking at ways to cut down on reporting requirements and paperwork, so that investigators can submit summaries of case reports rather than each individual document.

Trials that involve multiple sites also have to get approval from each institutions' review committee, referred to as an institutional review board. This can take months or even years.

To adapt to the multicentre climate, the US Office for Human Research Protections in Rockville, Maryland, which oversees human studies funded by the US National Institutes of Health, has proposed changes to its guidelines that would require designation of a single review board for each project. However, individual research centres may be reluctant to loosen their hold on the reins.

The clinical-trial system may be outdated, but at least it is not inflexible. Most radical designs have probably been tried in some capacity. Getting rid of control groups? It happens in rare-disease trials. Blending phases? Some trials are already starting to do this. Still, even small tweaks require a leap of faith when they are being made to a complex system with such high stakes — the protection of human health.

The way that clinical trials are designed and run may need to be revolutionized, but "there's not going to be a single solution because it's a multifactorial process," says Richard Schilsky, an oncologist at the University of Chicago, Illinois. "Everybody who is a stakeholder in the clinical-trial process has to contribute to the solutions." ■

Heidi Ledford writes for Nature from Boston.

AS THE MOLECULAR ERA ARRIVES, ADAPTIVE TRIALS WILL BECOME MORE AND MORE RELEVANT.

the approach to test combinations of drugs that target AKT and another protein, called mTOR (K. L. Pitter *et al.* *PLoS ONE* 6, e14545; 2011). The mouse trial allowed the researchers to try more concentrations than the human trial, and suggested new ways of screening patients for inclusion in future clinical trials. "The more complicated the trial, the more likely it is that the co-clinical strategy would be beneficial," Holland says.

In 2009, the NCI invested US\$4.2 million in Pandolfi's co-clinical trials in prostate and lung cancer. In them, mice receive the same therapies as people do before the start of experimental treatments. At The Jackson Laboratory, based in Bar Harbor, Maine, which has a large personalized medicine programme, researchers graft tumour cells taken from patients into mice. These mice can then

treatments they should be randomized to. But because the results are reanalysed each time a woman completes her treatment, the participants are more likely to receive a treatment that has worked in people with genetically similar tumours. If a treatment does not perform well in any patients, it is cut from the programme; other treatments can then be added to the mix.

The FDA and the EMA have encouraged drug developers to embrace these designs. In addition to high-profile trials such as I-SPY2, Berry says that he has designed trials running the whole gamut of disease, from diabetes to pandemic flu. "Virtually every pharma company is doing this," he says. But Getz disagrees and says that he's seen a more muted response to adaptive trials. "Our clinical-trial protocols have become too complex across the board," says Getz. "And adaptive designs add to the logistical complexity."

COMMENT

POLICY Why governments should measure well-being and promote the good life **p.532**

EPIGENETICS New tools to address old genomic questions **p.534**



MUSIC Icelandic singer Björk combines music and nature in an app **p.537**

FUNDING Universities should spend less on bureaucracy and more on researchers **p.538**

ILLUSTRATIONS BY JONATHAN BURTON



Fund people not projects

John P. A. Ioannidis proposes ways to save scientists from spending all their time writing grants.

The research funding system is broken: scientists don't have time for science any more. Because they are judged on the amount of money they bring to their institutions, writing, reviewing and administering grants absorb their efforts¹. The requirement that they promise taxpayers specific results to justify research tends to invite either exaggeration or boringly predictable projects. Yet the research behind 30% of the pivotal papers from Nobel laureates in medicine, physics and chemistry was done without direct funding².

Every scientist recognizes this problem

and hopes for a solution. Although detailed proposals may be indispensable for some projects, such as rigorous clinical trials and large-scale collaborative research, ideas abound for more efficient ways to fund general research. Some organizations are already experimenting. Multiple options could co-exist, with portions of the budget earmarked for different schemes.

Here are some of the most promising proposals to reduce the amount of time scientists spend trying to fund their research, and the pros and cons of each (see table). Definitive fixes would require major

system overhauls, which are likely to make some scientists justifiably nervous. But smaller, pilot efforts that enable us to evaluate what works could begin right away.

FUND EVERYBODY (OR A LUCKY FEW)

Some — or all — of the research budget could be allocated to eligible scientists in equal shares, or given to a few lucky ones at random. With egalitarian sharing, each scientist would receive only a small amount, which could quickly evaporate without returns when research costs are high. But scientists in some fields — mathematics, ►

► say — could achieve much on a small share; and in some settings the shares could be substantial. For instance, if half of the US\$31.2 billion that the US National Institutes of Health spends on research each year was shared equally among 300,000 researchers, each would get more than \$50,000 a year.

Lottery distribution, too, flies against the principle that science funding should be meritocratic. Still, some agencies are trying it — the Foundational Questions Institute in New York, which tackles key questions in physics and cosmology, uses a lottery system to award its mini-grants, which range in value from \$1,000 to \$15,000. Such an approach may not be as radical as it sounds: the imperfections of peer review mean that as many as one-third of current grants are effectively being awarded at random³. This situation will only worsen as falling acceptance rates encourage investigators to bombard agencies with proposals, leaving fewer qualified reviewers to judge each one. The downside of using aleatoric allocation is that not every deserving scientist will be funded.

FUND ACCORDING TO MERIT

Leading thinkers and experimenters worthy of unconditional support could be identified through peer assessment of their work and credentials. Appraisals of project-based proposals already take a scientist's merit into account, but they typically give less weight to it than to the project plan. Peer assessment does not work well for early-career scientists, who have a short track record. But for those more established in their field, a career trajectory offers a wealth of information. By contrast, an isolated project is only a snapshot.

In the MacArthur Fellows Program, for example, meticulous peer assessment is used to select 20–30 individuals a year on the basis of exceptional creativity and promise for important advances. Recipients do not have to justify what they do with the \$500,000 award, which is spread over 5 years. However, close scrutiny of an individual's career may become prohibitive for systems that award thousands of grants — it might save grant recipients time, but it adds to the administrative load of reviewers. The approach is also vulnerable to favouritism, in which only elite individuals and

“It is a scandal that billions of dollars are spent on research without knowing the best way to distribute that money.”

lines of research are selected and thousands of scientists doing quality, smaller-scale science are left out.

To avoid the subjectivity and burden of evaluating thousands of careers, an automated system to evaluate relative merit would have to be devised. Such a system would depend on objective indices. The share of the annual funding budget scientists receive would be based on their value, calculated with a pre-specified formula.

Metric-based appraisals are familiar to many scientists already, particularly those in European countries. The UK Research Assessment Exercise, for example, relies on them. It is a much hated and debated system for evaluating departments, but its replacement, the Research Excellence Framework,

will rely even more on indices when it comes into effect in 2014. Metrics also underlie many hiring decisions in Italy, a country that is struggling to remedy widespread nepotism, and in Germany's Max Planck institutes. However, most of these assessments are simplistic, focusing on number of peer-reviewed publications, or inappropriate — looking at the impact factor of the journal rather than of an individual article. More sophisticated formulae are needed if a scientist's merit is to be captured.

Furthermore, indices are open to gaming, although some are more difficult to influence than others. To counter this, the system could use indices that exclude self-citations and capture quality rather than quantity (such as average citations per paper instead of number of papers), discourage gift authorship by adjusting for co-authors and penalize quantity that is not accompanied by quality. Several metrics could be combined.

Funding systems could reward good scientific citizenship practices, such as data sharing⁴, high-quality methods, careful study design and meticulous reporting of scientific work⁵. Openness to collaboration, non-selective publication of 'negative' findings, balanced discussion of limitations in articles and high-quality contributions to peer-review, mentoring, blogging or database curation could also be encouraged. Researchers might be rewarded for publishing reproducible data, protocols and algorithms. However, some citizenship practices are difficult to capture in automated databases, so would be subject to the disadvantages of peer assessment.

OPTIONS FOR REVAMPING THE FUNDING SYSTEM

Option	Pros	Cons	Example	Who would be funded?
Egalitarian (fund everybody)	Avoids peer-review biases Gives sufficient amounts to scientists doing low-cost research Small administrative burden	Does not support large research efforts Does not recognize exceptional scientists	Some universities fund the salaries of all their faculty	All
Aleatoric (fund at random)	Avoids peer-review biases Small administrative burden	Will not capture all deserving scientists	Foundational Questions Institute	Flexible
Assessment of career	Captures career trajectory Has gold-standard status	Is vulnerable to favouritism Inappropriate for young researchers Is labour-intensive	MacArthur Fellows Program	Few elite scientists (or else administratively burdensome)
Automated impact indices	Eliminates favouritism Evaluates many applicants with ease Approaches objectivity	There are many indices, all with flaws; no consensus about best one to use Indices can be gamed Databases have shortcomings (such as imperfect citation coverage, entry errors, name disambiguation problems)	UK Research Excellence Framework	Flexible
Scientific citizenship	May improve science, if good practices are rewarded and bad ones penalized	Automation is not yet possible for data gathering, and is difficult for some citizenship practices Has peer-review biases	Financial incentives to peer reviewers	Could be extended to many scientists only for aspects that can be automated
Projects with broad goals	Proposals are easy to write and review Formulating work can be flexible Permits targeted innovation	Does not eliminate project proposals Is vulnerable to favouritism Holds potential for exaggerated promises and claims	NIH Director's Pioneer Awards Howard Hughes Medical Institute	Few elite scientists

Two or more options can also be combined (for example, automated impact indices plus evaluation of scientific citizenship).



STATE BROAD GOALS

Another way to save time would be to simplify the application. Researchers could be asked, for example, to submit short summaries of their intended research, describing broad goals only. Such applications require less effort to write, review and administer and would allow flexibility in carrying out the work, if funded. Examples include the NIH Director's Pioneer Awards and the Howard Hughes Medical Institute (HHMI) grants. The HHMI selects 300 established investigators and 50 young investigators through peer assessment of their credentials and of proposals for high-risk, uncertain prospects of innovation. Its awards are usually renewed after 5 years, requiring simple documentation of effort rather than demonstrated results, although results are needed for renewal after 10 years. Nevertheless, any system that demands high-risk innovative goals, and requires results, generates potential for exaggeration.

IGNORE GRANT PORTFOLIOS IN PROMOTIONS

Many institutions use the size of a scientist's grant portfolio as a basis for tenure and promotion. This practice may prompt scientists to prepare multiple grant applications focusing on expensive, even if dull, projects and to abandon brilliant ideas that need limited

funding to test. But going after many and expensive grants costs institutions money, because both scientists and administrators spend an inordinate amount of time evaluating proposals, supplements, timesheets and justifications, as well as progress and final reports.

Many large projects never result in a scientific achievement, so even if the strategy brings in short-term grant funding, it may not pay off in the long term. The size of a portfolio should therefore not be a criterion for promotion; committees should focus instead on real work and achievements. Judging scientists by the size of their portfolio is equivalent to judging art by how much money was spent on paint and brushes, rather than the quality of the paintings.

WHAT WE CAN DO NOW

All of the options above could be achieved — either through small, progressive steps, or through more extensive changes to the system. A major overhaul is likely to take years to implement, and will meet with wide resistance and debate. Smaller steps, such as changing promotion and tenure criteria, are easier to make. Pilot programmes of proposal-free or broad-goal-based funding can be incorporated into existing funding structures.

There are issues still to be resolved. All funding options face a tension over how many scientists should receive awards, and there is no good evidence on whether it is better to give fewer scientists more money or to distribute smaller amounts between more researchers. Some funding schemes are well suited to funding numerous scientists; others favour elitism (see table).

We will need to find ways to figure out which approach works best. It is a scandal that billions of dollars are spent on research without knowing the best way to distribute that money. Retrospective assessments are easy, but subject to confounding when comparing groups that were funded through different schemes. For example, one study has found that HHMI-funded investigators publish more high-impact papers and get more recognition than matched NIH-funded peers⁶. But it's impossible to match scientists perfectly: the prestige of the HHMI name alone may lead to more peer-based recognition. Prospective comparisons are more reliable, but require long follow-up. For example, controlled trials could randomize consenting scientists to different funding schemes, then compare surrogate metrics and long-term successes.

Ultimately, funding schemes should support the long-term goals of science. Few isolated research efforts have an immediate, substantial and durable impact; successful translation of basic research to practical applications occurs sparingly and with average delays of almost three decades⁷. The aim of science is to expand our knowledge base, which, eventually, yields useful applications. This is what scientists entered their profession to do, so requiring them to spend most of their time writing grants is irrational. It's time to seriously consider another approach. ■

John P. A. Ioannidis is the C. F. Rehnborg chair in disease prevention, professor of medicine and of health research and policy, and director of the Stanford Prevention Research Center at Stanford University School of Medicine, and professor of statistics at Stanford University School of Humanities and Sciences, Stanford CA 94305. e-mail: jioannid@stanford.edu

1. Editorial. *Sci. Am.* **304**, 10 (2011).
2. Tatsioni, A., Vava, E. & Ioannidis, J. P. *FASEB J.* **24**, 1335–1339 (2010).
3. Graves, N., Barnett, A. G. & Clarke, P. *Nature* **469**, 299 (2011).
4. Baggerly, K. *Nature* **467**, 401 (2010).
5. Simera, I., Moher, D., Hoey, J., Schulz, K. F. & Altman, D. G. *Eur. J. Clin. Invest.* **40**, 35–53 (2010).
6. Azoulay, P., Graff Zivin, J.S., & Manso, G. *Incentives and Creativity: Evidence from the Academic Life Sciences* (NBER Working Paper No. 15466, 2009).
7. Contopoulos-Ioannidis, D. G., Alexiou, G. A., Gouvias, T. C. & Ioannidis, J. P. A. *Science* **321**, 1298–1299 (2008).



E. MARTINO/PANOS

Family support in Brazil bolsters well-being even when economic resources are limited.

Time to legislate for the good life

Charles Seaford argues that a clear measure of well-being should be devised to help people judge how government policies affect their quality of life.

In a speech in May, British prime minister David Cameron remarked that a programme under way at the UK Office for National Statistics (ONS) to collect social data and develop decision-making rules based on well-being “may be the most quietly radical thing this government is doing”. The riots in the United Kingdom last month — an expression of social dissatisfaction that was more than a response to economic conditions — highlight the importance of this approach.

Governments in France, Italy, Germany, Spain, the United States, Australia, China and Ecuador are also taking steps to measure quality of life as well as economic growth. The European Commission and the Organisation for Economic Co-operation and Development are harmonizing these efforts. And the General Assembly of the United Nations has called upon member states to place greater emphasis on measures of well-being in their evaluations of social and economic development. My colleagues and I at the New Economics Foundation (NEF) in London have been arguing for such a focus since 2001; we are now working with the UK government and the European Commission to ensure that these efforts are translated into practical implications for public policy.

This interest in well-being — and its subjective measurement — is good news. Economic

growth is just one of many tools for bringing about good lives. Political decisions involve trade-offs — between, say, fostering economic growth and stable communities, or agreeable urban landscapes. The traditional focus on gross domestic product (GDP) as a target biases these decisions. The result is lower levels of public well-being than could be the case if people’s quality of life was the priority. As economic activity places a greater strain on the environment than many other routes to happiness — such as spending time with one’s family — this bias is also bad for sustainability (see ‘Good lives needn’t cost the Earth’).

There are two key challenges for researchers, politicians and policy-makers: first, to gather and interpret new data, so as to create

a much fuller science of well-being to rival traditional economics; and second, to create public understanding of some headline measure of well-being and of the role of policy in influencing it, in order to create the political will to use the new science.

THE EVIDENCE

Well-being is variously defined. Psychologists see it as ‘good functioning’ or the meeting of psychological needs¹, an approach that emphasizes relationships, autonomy, competence and purpose. Economists use more abstract terms such as ‘happiness’ or ‘utility’.

Social surveys over several decades have shown that economic and social policies affect aspects of well-being, however it is defined. Income correlates with well-being, but only up to a certain level, which varies between countries. In the United States, for example, earnings above US\$75,000 don’t add much more happiness². Studies also reveal that loss of income is more damaging than a gain is beneficial, and unemployment is more damaging to well-being than is the consequent loss of income. Casual employment is bad for well-being, but self-employment is good, at least for those earning a decent income. Commuting is bad.

Moderate economic growth is good for well-being, but the disruption to lives that is

SUMMARY

- Governments worldwide should measure and monitor well-being using national surveys.
- An internationally accepted well-being statistic should be developed and communicated to the public.
- Public debate around this statistic will prompt policy-makers to maximize well-being over economic growth.

associated with a rapid boom is bad. Inflation is harmful — but a 1% rise in inflation hurts less on average than a 1% rise in unemployment. Fluctuating rates of inflation are worse than stable ones. High levels of short-term debt are associated with low values of well-being. Effective democratic structures, social trust, social networks and time spent socializing are all important for well-being, as are access to parks and views of greenery for city-dwellers.

Levels of well-being vary sharply between countries. Sub-Saharan Africa and Eastern European countries score badly. Scandinavia and Latin America do well. In general, there is a positive association between social equality and well-being, although there are exceptions: inequality is high in Latin America, for example, but family support plays a strong role in social security³.

So an economic policy driven by what we already know about well-being would place more emphasis than do current economically centred policies on securing full employment, on stability, and on preserving and strengthening communities and ensuring a reasonable level of income for all. Governments have not yet made those changes. What will cause them to do so?

A SINGLE MEASURE

A headline statistic that captures well-being — a measure that voters recognize and can use to hold politicians to account, like the unemployment or economic growth figures — could make the difference. Unless voters care, politicians won't care. And voters won't care without a number — a measure that tells them how things have been going and might be projected to the future. The GDP statistic has had this effect since it was invented in 1934.

The Stiglitz commission, set up in 2008 by French president Nicolas Sarkozy, has laid

some of the groundwork for developing a measure of well-being⁴, but there is more to do. The measure must be able to move and be seen to move over time, in the way that GDP figures do. It must be in a format that allows international comparisons, perhaps accounting for intercultural differences by focusing on movements rather than absolute values.

Most important, it must link well-being to social and economic conditions — and that will require further research. The survey results published are rarely linked to particular policy decisions. And if they are not 'official', the statistics may go unnoticed beyond a given field. The causal factors underlying correlations between happiness and social conditions are often poorly understood, the variables are difficult to control for and the results can seem inconsistent between studies. Large-scale studies that investigate many potential drivers of well-being and also include adequate controls are scarce.

The United Kingdom is leading the way to better measurement with an annual survey by the ONS (on which the NEF has advised) that will probe the well-being and circumstances of 200,000 people. Since April, the ONS has included in its Integrated Household Survey four subjective questions: how satisfied people are with their lives; how happy they were yesterday; how anxious they were yesterday; and how worthwhile they think the things they do are. The survey also asks a large number of questions about objective circumstances that are influenced by government policy — for example, health, housing, education, household income and employment patterns and benefit entitlements.

With such information it will become possible to trace statistical relationships and thereby to assess the relative impact of various policies on well-being. Well-being will then become the common currency for a

new form of cost-benefit analysis, which will not replace political judgement and bargaining, but should inform it. Policy-makers will be able to ask, for example, what have been the relative impacts of steps to reduce unemployment in an area and steps to preserve the environment? Have attempts to increase community cohesion and increase economic activity worked? Have the public-health measures in one city been more effective at increasing well-being than those in another?

THEORY IN ACTION

Our recommendation to the ONS, and to other governments, is that the main measure publicized each year should be based on perceived levels of well-being rather than the drivers of it. That is, it should be based on the responses to the four subjective questions asked in its household survey, rather than on the objective questions and the resulting data. It should take the form of a percentage of the population above a certain threshold level of well-being — the percentage that is flourishing.

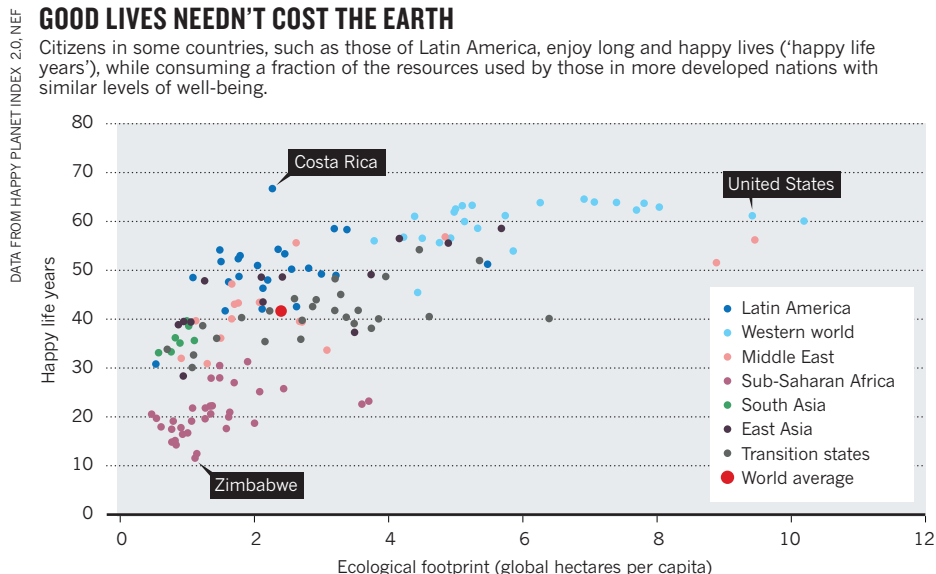
The analytical work identifying the drivers of this headline measure must also be presented, in a form that can be picked up by the media and can inform a public debate about which policies to pursue. Voters will come to see that movements in the measure — like those of GDP or inflation or carbon emissions — are the responsibility of politicians. Politicians will then have to take the well-being measures seriously, and act on the advice of the analysts.

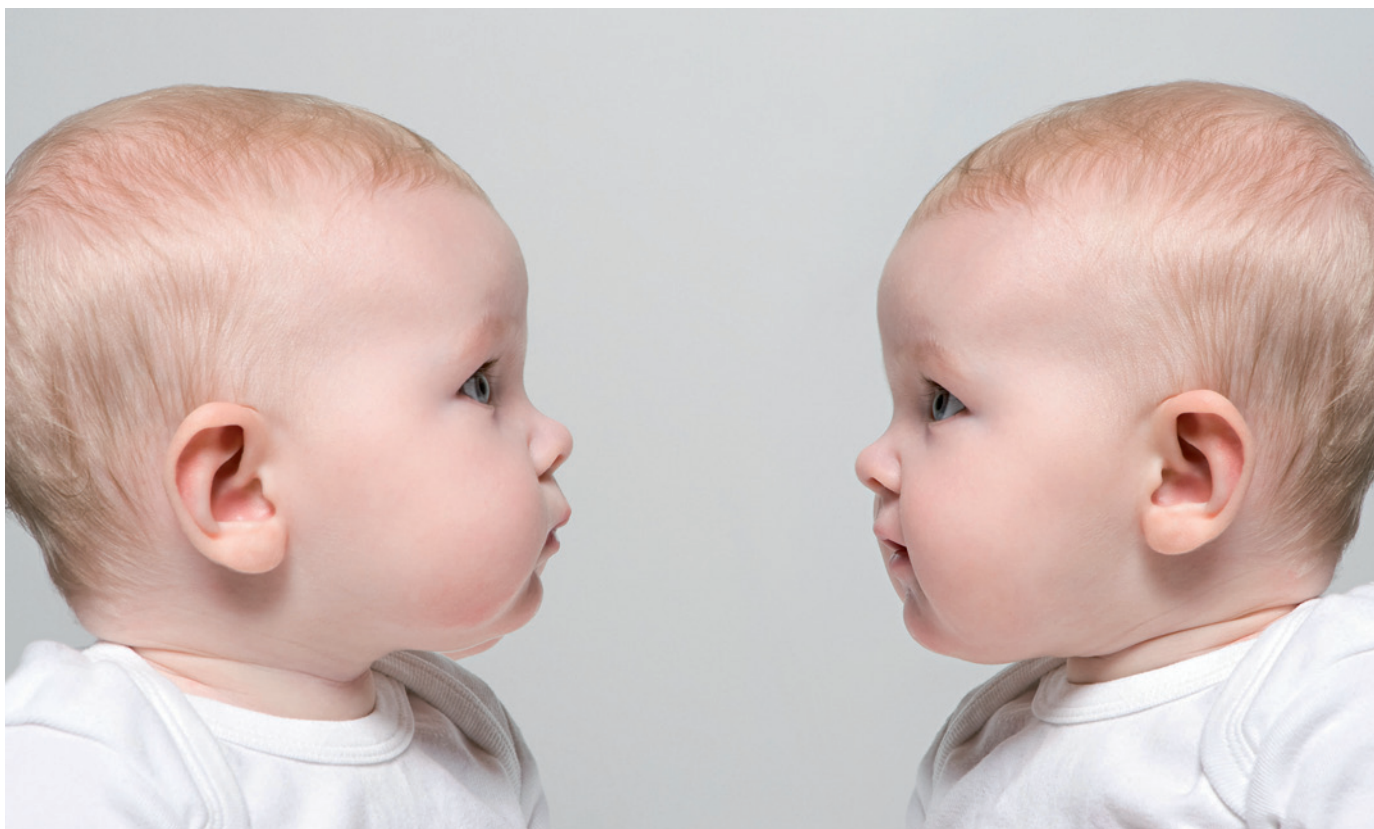
Such discussion must not be conducted behind closed ministry doors. The statistical authorities should make the data publicly available at a level of granularity (including geographical detail) useful to a wide range of researchers. To probe causality, researchers will need funding for field experiments and qualitative work. And somehow, politicians and officials will need to be made to listen to the results even when they do not confirm existing views.

All of this is possible, but it will not happen automatically. We must continue to push. ■

Charles Seaford is head of the Centre for Well-being at the New Economics Foundation, 3 Jonathan Street, London SE11 5NH. He is a member of the UK National Statistician's Advisory Forum on Well-being.
e-mail: Charles.Seaford@neweconomics.org

1. Ryan, R. M., Huta, V. & Deci, E. L. *J. Happiness Stud.* **9**, 139–170 (2008).
2. Kahneman, D. & Deaton, A. *Proc. Natl Acad. Sci. USA* **107**, 16489–16493 (2010).
3. New Economics Foundation *Well-being Evidence for Policy: An Annual Review* (in the press, 2011); available at www.neweconomics.org
4. *Report by the Commission on the Measurement of Economic Performance and Social Progress* (2009); available at go.nature.com/fnl7he





PHOTOLIBRARY

Inherited and environmental factors can switch genes on and off to give each of us, including monozygotic twins, a unique identity.

EPIGENETICS

Beyond face value

Jonathan Weitzman relishes two accounts of how environment can influence the script of our genome.

As an identical twin, I have always been fascinated by what determines who we are. Nature's clones are never truly identical, so what explains the differences between my brother and myself? How much of our identity is inherited; how much acquired by interacting with the environment? The field of epigenetics, standing at the interface between our environment and our genes, is beginning to offer answers.

Epigenetics explores how genetically identical entities, whether cells or whole organisms, display different characteristics, and how these are inherited. The past century witnessed amazing advances in our understanding of genetics, but secrets remain hidden within the genome. Epigenetics research is now blossoming, offering a potential panacea for these post-genome blues.

Two timely books open up this emergent field: *Epigenetics* by Richard Francis and *The Epigenetics Revolution* by Nessa

Epigenetics: The Ultimate Mystery of Inheritance

RICHARD C. FRANCIS

W. W. Norton: 2011. 224 pp. £19.99, \$25.95

The Epigenetics Revolution: How Modern Biology is Rewriting Our Understanding of Genetics, Disease and Inheritance

NESSA CAREY

Icon Books: 2011. 320 pp. £17.99

Carey offer very different takes. Francis's thoughtful and succinct book focuses on the narrative and the excitement of discovery, rather than on the nitty-gritty details at the molecular level. His personal tour includes anecdotes from his travels around the world and allusions to popular culture.

Carey's book is more DNA-centric, focusing on epigenetic mechanisms and the chemistry of chromatin, which defines how DNA is packaged around proteins in the nucleus. Her book combines an easy style with a textbook's thoroughness.

Conveying epigenetics to a broad readership presents challenges. The field is in flux and the concepts are abstract and technically demanding. Researchers struggle to come up with a unanimously accepted definition of the term epigenetics. And there is confusion in the literature between epigenetic phenomena and epigenetic mechanisms. The first describes the transmission of heritable characteristics, thereby resembling genetics; the second focuses on how genes are regulated, and is more aligned with molecular biology. Francis is interested in the modes of inheritance and how these are affected by environmental events. Carey refers more to the mechanistic events that change how genes are switched on and off without altering the genes themselves.

Epigenetics, for Francis, is a form of cellular control over

[NATURE.COM](http://go.nature.com/ugle41)

For Nature's special on the Human Genome at Ten: go.nature.com/ugle41

gene activity. He presents the cell as a theatrical production, noting that the standard view — with genes as directors, proteins as actors and other biochemicals as stagehands — needs to be reassessed. As an alternative to the ‘executive gene’ idea, in which the genome is in control, he urges readers to think about the ‘executive cell’. Francis relegates genes to members of the cast, and insists that they are regulated by, not regulators of, cellular states.

A seasoned science writer, Francis is at his best when describing abstract concepts in a historical context. A chapter on developmental biology reminds us that the idea that form emerges gradually during development (known as epigenesis) arose in

“What makes the field so exciting is that we have new tools to address old questions.”

response to an earlier concept, which held that form is predetermined from the beginning (preformationism). Francis cautions us not to fall into the trap of using the ‘genetic blueprint’

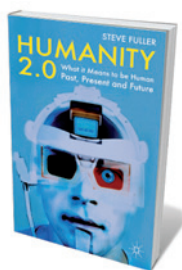
or ‘genetic program’ metaphors, which he sees as contemporary preformationism or genomic determinism. He ends with an interesting concept — the ‘Janus gene’ — which has both the outward causal aspect of gene function and an inward aspect that is responsive to external cues.

Carey’s book is full of illustrations and entertaining metaphors. She describes the nucleosome as DNA wrapped around “eight ping-pong balls ... like a long liquorice whip around marshmallows”, and sees DNA as a film script, with plenty of room for interpretation and retakes. Carey’s experience of the biotechnology industry shows in her concluding remarks on the pros and cons of our growing understanding of epigenetics for drug discovery, and on understanding the impact of diet and environment on disease.

These two bold attempts to bring epigenetics to a wide audience are at the vanguard: more books will follow as the field matures. But future authors should avoid using the words ‘mystery’ and ‘revolution’ in their titles. What makes the field so exciting is that we have new tools to address old questions. Unravelling life’s mysteries? That’s not revolution, it is why we become scientists in the first place. My twin brother and I are both professors studying genetics — I still wonder whether that choice was determined by genetics or the environment. ■

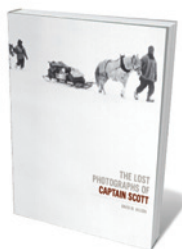
Jonathan Weitzman is professor of genetics at the Université Paris Diderot and director of the CNRS unit for Epigenetics and Cell Fate, Paris, France.
e-mail: jonathan.weitzman@univ-paris-diderot.fr

Books in brief



Humanity 2.0: What it Means to be Human Past, Present and Future
Steve Fuller PALGRAVE MACMILLAN 288 pp. £19.99 (2011)

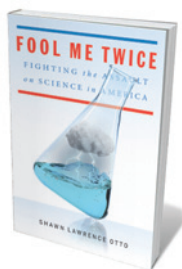
Forget *Frankenstein* and *Metropolis*, says sociologist Steve Fuller. Neuroscience and technology are on a collision course that will catapult us to ‘transhumanity’, a future in which the state of being human starts to look decidedly alien. Fuller works his way through science, policy, history and philosophy, covering human biological roots and our aspirations to transcend them; the possibility of enhanced evolution as technologies converge; a ‘new theology’ inspired by figures such as Jesuit palaeontologist–philosopher Pierre Teilhard de Chardin; and our necessary quest for moral horizons.



The Lost Photographs of Captain Scott

David M. Wilson LITTLE, BROWN 192 pp. \$35 (2011)

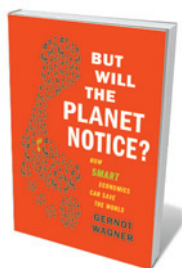
‘Lost’ is a refrain commonly associated with Antarctic explorer Robert Falcon Scott — lost race, lost hopes, lost lives. Now, a stunning record of his last months has been found, giving a glimpse of fieldwork at the ‘final frontier’. In keeping with the research aims of the expedition, Scott shot 120 photographs between Cape Evans and Beardmore Glacier before his team’s final, ill-fated push. Most of the images — icy panoramas, portraits, snaps taken on the move — are published here for the first time, with text by polar historian David M. Wilson, great-nephew of Scott’s science chief Edward Wilson.



Fool Me Twice: Fighting the Assault on Science in America

Shawn Lawrence Otto RODALE 384 pp. \$25.99 (2011)

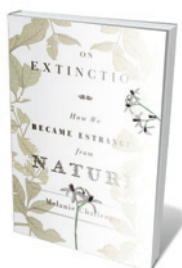
Less than 2% of US senators and representatives have a scientific background, but more than 40% are lawyers. Shawn Lawrence Otto suggests that is why US policy-making is rich in rhetoric and poor in reasoning. Reminding us of the crucial separation of church and state at the birth of the United States, Otto traces shifts in national attitudes to science and technology — from early wonder through atomic-era fear to widespread rejection. That these shifts are happening precisely when big science is needed to tackle global challenges should, he says, push researchers to re-engage with politicians.



But Will the Planet Notice? How Smart Economics Can Save the World

Gernot Wagner HILL AND WANG 272 pp. \$27 (2011)

The plastic-bag militants have got it wrong, says economist Gernot Wagner. We can save the planet by cutting out scientists, politicians and environmentalists in favour of economists. It’s they, he claims, who have revealed our mixed-up thinking, as we buy locally grown apples chilled for months, or clear rainforests to grow soya. Wagner calls for smart economics, such as capping pollution and paying for free resources. Billions of us must be motivated for change to be meaningful — and the best incentive is market forces, he says.



On Extinction: How We Became Estranged from Nature

Melanie Challenger GRANTA 272 pp. £20 (2011)

Writer and poet Melanie Challenger explores extinction, both biological and cultural. Roving through wild landscapes — Antarctica and the Arctic, disused Cornish tin mines and old whaling stations in South Georgia — she interweaves the science on ecological loss and climate change with histories of failed industries, extinguished languages and wars. From albatrosses on Bird Island, South Georgia, to the Inuit of Canada’s Frobisher Bay, Challenger combines her meditations on our fragmenting world into a finely integrated study of loss.

gene activity. He presents the cell as a theatrical production, noting that the standard view — with genes as directors, proteins as actors and other biochemicals as stagehands — needs to be reassessed. As an alternative to the ‘executive gene’ idea, in which the genome is in control, he urges readers to think about the ‘executive cell’. Francis relegates genes to members of the cast, and insists that they are regulated by, not regulators of, cellular states.

A seasoned science writer, Francis is at his best when describing abstract concepts in a historical context. A chapter on developmental biology reminds us that the idea that form emerges gradually during development (known as epigenesis) arose in

“What makes the field so exciting is that we have new tools to address old questions.”

response to an earlier concept, which held that form is predetermined from the beginning (preformationism). Francis cautions us not to fall into the trap of using the ‘genetic blueprint’

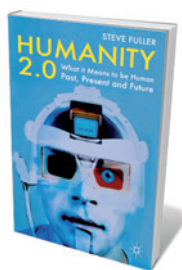
or ‘genetic program’ metaphors, which he sees as contemporary preformationism or genomic determinism. He ends with an interesting concept — the ‘Janus gene’ — which has both the outward causal aspect of gene function and an inward aspect that is responsive to external cues.

Carey’s book is full of illustrations and entertaining metaphors. She describes the nucleosome as DNA wrapped around “eight ping-pong balls ... like a long liquorice whip around marshmallows”, and sees DNA as a film script, with plenty of room for interpretation and retakes. Carey’s experience of the biotechnology industry shows in her concluding remarks on the pros and cons of our growing understanding of epigenetics for drug discovery, and on understanding the impact of diet and environment on disease.

These two bold attempts to bring epigenetics to a wide audience are at the vanguard: more books will follow as the field matures. But future authors should avoid using the words ‘mystery’ and ‘revolution’ in their titles. What makes the field so exciting is that we have new tools to address old questions. Unravelling life’s mysteries? That’s not revolution, it is why we become scientists in the first place. My twin brother and I are both professors studying genetics — I still wonder whether that choice was determined by genetics or the environment. ■

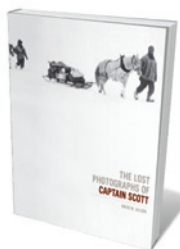
Jonathan Weitzman is professor of genetics at the Université Paris Diderot and director of the CNRS unit for Epigenetics and Cell Fate, Paris, France.
e-mail: jonathan.weitzman@univ-paris-diderot.fr

Books in brief



Humanity 2.0: What it Means to be Human Past, Present and Future
Steve Fuller PALGRAVE MACMILLAN 288 pp. £19.99 (2011)

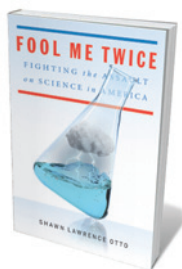
Forget *Frankenstein* and *Metropolis*, says sociologist Steve Fuller. Neuroscience and technology are on a collision course that will catapult us to ‘transhumanity’, a future in which the state of being human starts to look decidedly alien. Fuller works his way through science, policy, history and philosophy, covering human biological roots and our aspirations to transcend them; the possibility of enhanced evolution as technologies converge; a ‘new theology’ inspired by figures such as Jesuit palaeontologist–philosopher Pierre Teilhard de Chardin; and our necessary quest for moral horizons.



The Lost Photographs of Captain Scott

David M. Wilson LITTLE, BROWN 192 pp. \$35 (2011)

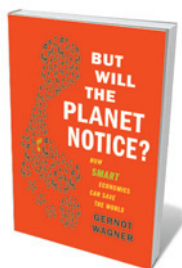
‘Lost’ is a refrain commonly associated with Antarctic explorer Robert Falcon Scott — lost race, lost hopes, lost lives. Now, a stunning record of his last months has been found, giving a glimpse of fieldwork at the ‘final frontier’. In keeping with the research aims of the expedition, Scott shot 120 photographs between Cape Evans and Beardmore Glacier before his team’s final, ill-fated push. Most of the images — icy panoramas, portraits, snaps taken on the move — are published here for the first time, with text by polar historian David M. Wilson, great-nephew of Scott’s science chief Edward Wilson.



Fool Me Twice: Fighting the Assault on Science in America

Shawn Lawrence Otto RODALE 384 pp. \$25.99 (2011)

Less than 2% of US senators and representatives have a scientific background, but more than 40% are lawyers. Shawn Lawrence Otto suggests that is why US policy-making is rich in rhetoric and poor in reasoning. Reminding us of the crucial separation of church and state at the birth of the United States, Otto traces shifts in national attitudes to science and technology — from early wonder through atomic-era fear to widespread rejection. That these shifts are happening precisely when big science is needed to tackle global challenges should, he says, push researchers to re-engage with politicians.



But Will the Planet Notice? How Smart Economics Can Save the World

Gernot Wagner HILL AND WANG 272 pp. \$27 (2011)

The plastic-bag militants have got it wrong, says economist Gernot Wagner. We can save the planet by cutting out scientists, politicians and environmentalists in favour of economists. It’s they, he claims, who have revealed our mixed-up thinking, as we buy locally grown apples chilled for months, or clear rainforests to grow soya. Wagner calls for smart economics, such as capping pollution and paying for free resources. Billions of us must be motivated for change to be meaningful — and the best incentive is market forces, he says.



On Extinction: How We Became Estranged from Nature

Melanie Challenger GRANTA 272 pp. £20 (2011)

Writer and poet Melanie Challenger explores extinction, both biological and cultural. Roving through wild landscapes — Antarctica and the Arctic, disused Cornish tin mines and old whaling stations in South Georgia — she interweaves the science on ecological loss and climate change with histories of failed industries, extinguished languages and wars. From albatrosses on Bird Island, South Georgia, to the Inuit of Canada’s Frobisher Bay, Challenger combines her meditations on our fragmenting world into a finely integrated study of loss.

Tuning into translation

Ellen Bialystok relishes a journey across the porous borders between languages.

If you have ever despaired about the possibility of meaningful communication, *Is That a Fish in Your Ear?* may reignite your anxiety.

British translator David Bellos begins his journey through the fascinating history and convoluted present of translation with questions that seem simple, but swiftly enter complex territory. Language, he tells us, is a malleable thing. The boundaries we naively draw between languages are permeable, and all communication is interpretation.

Translation encourages us to reconsider such issues as culture, nationalism, ethics and, above all, the part played by the mind — the precision and artistry that make translation a creative process. Yet it is the mind, ironically, that is given the least attention in this otherwise wide-ranging description of translation.

Bellos takes us through the meaning of the word 'translation' in different languages and times; the special case of biblical translation and its influence on religious following; translation in a modern industrial world where products are distributed internationally; and technological innovations that increasingly assign the task of translation to computers.

Bellos explains why even the designation of your mother tongue is not a simple judgement. The language of home can be supplanted, for instance, by that of education as the 'most native' of one's languages. Neither is the choice of source and receiving languages for a translation straightforward, as exemplified by the French passages in the original 1869 Russian version of Tolstoy's *War and Peace* that are sometimes translated, sometimes not. Add to this the widespread absorption of words from one language (particularly English) into another, together with the changes in language usage arising from texting and tweeting, and the very concept of a fixed language dissolves.

In the absence of set structure and stable reference, how is translation possible at all? To answer this conundrum, Bellos ranges across linguistic theory, historical linguistics, personal experience and a deep knowledge of literature. Three of his examples shed light on the process of translation by describing the practical basics of the task, the politics that surround language choices and the systems used to get the job done.

First, to understand the challenge of translation fully, we must look to poetry, the

subtlety of literary forms. A good translation must be a good match, but which of a poem's many dimensions need to be preserved? Meaning? Form? Metre? Rhyme, assonance or alliteration?

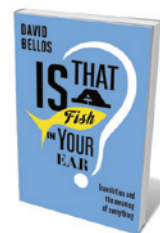
Second, decisions on which languages get translated into which others, and which social style of speech is used, are inherently political. Only a small number of the world's languages ever get translated, and an even smaller number make up the receiving languages; English wins again here. Bellos writes of the belief "that only some languages were suited to civilized thought" — a relic of nationalism and colonialism that continues to exert an insidious influence.

Finally, there is the business of translation. How does the United Nations, for instance, provide simultaneous interpretations of proceedings for its entire assembly? The answer is an ingenious system that combines technology with economics, and overcomes the complexities of the task by deploying limited numbers of translators on strict rotas.

Bellos has little to say about how



The complexities of simultaneous translation are a challenge at United Nations assemblies.



Is that a Fish in Your Ear? Translation and the Meaning of Everything
DAVID BELLOS
Faber & Faber: 2011.
384 pp. \$26, £20

languages are learned, but makes several allusions to a critical period in childhood for learning a second language. Although some theorists continue to endorse that notion, it is incompatible with the book's dynamic view of language and the latest research. In my view, Bellos is incorrect in his claim that no child develops lan-

guage without being exposed to it. In 1977, a community of deaf children in Nicaragua with no exposure to language spontaneously created Nicaraguan Sign Language, a perfectly systematic and rich system. A more detailed consideration of the role of the mind in language and translation would have been welcome.

Concepts of language have evolved from twentieth-century structuralist and generativist theories — which explain language in terms of grammar — into one in which language seamlessly integrates into the cognitive, social and emotional lives of the humans who use it. Almost 40 years ago, psychologist Roger Brown coined the term 'rich interpretation', describing the need to consider both what a child is doing and what he or she is saying in order to understand what the child is trying to convey. Although Bellos makes no reference to Brown, he implicitly extends Brown's idea, arguing that all meanings are embedded in dimensions outside language, and that a good translation captures many of them. He correctly identifies the central role of the mind in the process of interpretation that moves information between languages.

So don't despair: human communication is possible, as is the translation of meaning across the shifting boundaries of language, even though the translation and the original are never identical. Any translation can signify a rose; an excellent one will convey its fragrance. ■

Ellen Bialystok is a professor of psychology specializing in bilingualism at York University, Toronto, Canada.
e-mail: ellenb@yorku.ca



Q&A Björk

Digital diva

The Icelandic singer Björk's album *Biophilia*, set for release on 10 October, features songs on DNA, crystals, viruses and electricity — each accompanied by an iPad app that doubles as a music lesson. She explains how her upbringing and environmental activism has inspired her, and how touch-screen technology has changed the way she composes.

What is the *Biophilia* project?

It's a multimedia exploration of music, nature and technology, which is also educational. Each song has a linked iPad app — with animation, music scores and lyrics — that teaches kids the basics of music, such as rhythms and scales. I felt that the years between five and eight, when a child's brain is soaking up languages and learning to read and write, are the perfect time to absorb musical theory. And now I can become a music teacher. We are planning live shows and workshops for kids at science museums, including ones in San Francisco and Tokyo.

How did you put the tracks together?

I took algorithms of patterns in nature, put them into an iPad, played with them and then plugged the result into an instrument. The arpeggios on the bassline of one track, 'Thunderbolt', were created from the electrical discharge of a Tesla coil. A choppy, complex piece became 'Crystalline', which is about structure. 'Virus' is a kind of love song, where viruses take over; it is about arrangement. 'Hollow', a song about DNA, is about rhythm; 'Dark Matter' is about musical scales.

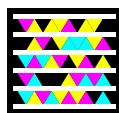
What inspired you to link nature and music?

I've always written songs while walking — in accompaniment with nature. I'd walk up a hill, and that would be the verse; when I'd reach the top, that would be the chorus. And if it was snowing, or sunny, that would result in very different songs. I would mould the melodies by singing them over and over. Afterwards I'd get the computer out. But touch screens have revolutionized the way I work: now I can write and arrange songs at the same time.

How else have touch screens influenced you?

I can compose more organically. Electronic music has been criticized for being too cold or grid-like. But with a touch screen the grid becomes like water. In the past I didn't have the right tools: on my album *Homogenic* I took electronic beats and distorted them to be like eruptions or lava — to be Icelandic. By using a touch screen, *Biophilia* has

NATURE.COM
For more on Björk's
Biophilia:
go.nature.com/aepg49



Scan the tag above with the free app from gettag.mobi

taken me to a different place. They are also good for education. The wrong way to be taught is with a book: it should be something you can turn in circles, and play with.

Is Iceland a big influence?

When I came up with the idea for *Biophilia* I was involved in an environmental battle to stop aluminium smelters from being built in Iceland. Then the bank crash happened and we were wondering how to create jobs. In Iceland — a volcanic place — a lot of people choose science subjects at university, especially geology, biology, physics and mathematics. But now, in a nation of 350,000, there are hundreds of unemployed biotech people. We're in trouble because we're not seeing the obvious. We need to find solutions to problems like pollution and energy that are sustainable, and I want to help.

Are you interested in science?

I was a maths nerd in school. And I was keen to learn more for this album. I watched a lot of educational DVDs, including ones on string theory. Some scientists are now saying that the Big Bang didn't exist. I think it is refreshing that the twenty-first century already has different 'creation myths' from the twentieth century. One track on *Biophilia*, 'Cosmogony', is about creation myths. The first verse is from Native American culture; the second, Sanskrit scriptures; the third, Australian Aboriginal culture; and the fourth, Big Bang theory. In a way, they are all saying the same thing.

Did you discuss the project with scientists?

I spoke to Oliver Sacks and read his book *Musicophilia*, on music and the brain. It was an inspiration for this project — and its title. As English isn't my mother tongue, I thought *biophilia* meant 'empathy with nature'. Then somebody told me it means 'love of life', which isn't far off. Sacks's book opens with a man who got hit by lightning and started hearing music in his head. He quit his job, learned the piano and became a composer. My favourite chapter is about the connection between Tourette's syndrome and jazz drumming.

What links science and music?

String theory is about strings that vibrate, which is essentially music. I've also read about the history of the music of the spheres, from Pythagoras to our time. Every hundred years or so, somebody tries to write the music of the spheres. It's always a different time or country, but it's always the same idea. Pythagoras was tapping into harmonics and how sound works, but that is similar to how atoms and the Solar System work. String theory is trying to unite the Solar System and the atoms. Like sound, you throw something out at the world and it comes back at you. ■

INTERVIEW BY ANDREW MITCHINSON

Correspondence

Bureaucracy savings should go to research

We believe that there is a pressing need to cut the cost of bureaucracy in universities and to channel those savings into research.

In Australia, for example, spending by the top ten research universities rose from almost Aus\$6 billion in 2003 (US\$3.9 billion at the time) to more than Aus\$10 billion in 2010 (2010 US\$9.2 billion), but only 30% of this was allocated to the employment of academics. This is worrying, given that a university's main objectives are research and teaching. Almost as much (26%) was spent on non-academic salaries; the rest went on assets such as branding, buildings and equipment. Similar spending patterns can be found in UK and US universities.

As well as the restrictive effect on academic appointments, this bureaucratic burden adversely affects scholarly output, with academics spending too much time on paperwork and internal meetings. Some must even forgo research altogether as they become swamped by administrative tasks.

Senior university officials should make radical changes to the spending habits of institutions, despite the likely resistance that would come from non-academics. Australian academic Alan Trounson, president of the California Institute for Regenerative Medicine (CIRM) in San Francisco, has shown that such changes are feasible. In 2009, he pledged to spend less than 6% of CIRM's revenues on administration costs, a figure that is better than for most commercial firms (see go.nature.com/ummgbk). As of February 2011, CIRM had spent 4.1% on administration (see go.nature.com/ntv5he).

Administration should enable universities to run efficiently, but it must be streamlined. A bold university that reforms its cost

structures will have more money to spend on its core business.

Nicholas Graves, Adrian G. Barnett *Queensland University of Technology, Queensland, Australia.*
n.graves@qut.edu.au

Ireland: a strong knowledge economy

I take issue with Mojgan Naghavi's and Derek Walsh's critique of Science Foundation Ireland (SFI; *Nature* **476**, 399; 2011). Ireland has moved steadily up the country rankings of research performance and is now in the top 20, according to Thomson Reuters Essential Science Indicators.

This success is largely a result of the rigour of the international competitive review process implemented by the SFI since 2000, intended to select only the best science for funding.

This strategy has been purposefully carried out by successive governments against a worsening domestic and international economic backdrop, exemplifying a consistent, considered and long-term approach to research in Ireland.

Also, the SFI does not directly recruit researchers. Rather, it allows researchers and host institutions to compete for research funding.

The foundation has made a significant investment in research infrastructure over the past decade. However, more than 75% of state investment has been, and continues to be, in the scientists themselves. The budget of the SFI in 2011, far from fading, was increased by 7% at a time when most public spending has been significantly curtailed.

Naghavi and Walsh's suggestion that Ireland is among nations with "ill-planned knowledge-economy ventures" is at odds with its impressive statistics on inward investment in research and development,

and the role of this in enhancing exports, increasing employment and adding value to Irish firms (see, for example, www.forfas.ie and www.idaireland.com).

Graham Love *Science Foundation Ireland, Dublin, Ireland.*
graham.love@sfi.ie

Ireland: global links pay off

I contend that Mojgan Naghavi's and Derek Walsh's claim of cronyism in the Irish university system is unfounded (*Nature* **476**, 399; 2011).

Excellence in research and teaching are the criteria for recruitment. The openness of our recruitment system is evident: more than 35% of doctoral candidates and 35% of postdoctoral researchers in Ireland are from overseas, and 26–42% of permanent academic staff across the seven Irish universities are international. They add strength to the cohort of tenured senior investigators, which is essential to the health of any advanced research system.

Ireland is building up its research base through a long-term plan to attract and retain the best talent and to develop its research infrastructure. Progress began in the 1990s as a result of investment from the European Union's Structural Funds and Framework Programmes for Research and Technological Development, which led to more international collaborations. Such investments are indicative of genuine strategic intent by Ireland's government and universities.

National policy initiatives have emphasized the need to invest in research as an integral part of economic development and, more recently, recovery (see, for example, go.nature.com/kwfpct). These have been implemented by consistent state and philanthropic investments in research, and by the ability of Irish universities to compete

successfully for international research funds. As a result, the country now has state-of-the-art laboratories and research centres (see, for example, go.nature.com/qzfaed) and, crucially, talented scientists to staff them.

Conor O'Carroll *Irish Universities Association, Dublin, Ireland.*
conor.ocarroll@iua.ie

More freedom for Turkish science

The future of science in Turkey is being undermined by recent government actions (*Nature* **477**, 131; 2011). Academic autonomy is crucial if Turkish scientists are to flourish at home and abroad. A partisan approach will not advance Turkey's society or stop the brain drain.

The government should take advantage of the Turkish economy's latest positive indications and increase the country's modernization assets. At present, just 0.6% of gross domestic product is invested in research and development.

To improve research morale, Turkey's government needs to reverse the centralization of science policy-making and give more power to scientific committees. These should be made up of competent researchers, who understand their own needs better than officials. It should reform scientific councils to allow them to operate democratically and with transparency, and abandon partisan legislation.

These measures would improve the credibility of science in society and help to educate a fatalistic and irrational public. Albert Einstein pointed out that imagination is more important than knowledge: the Turkish scientific community is simply seeking the prerogative and freedom to imagine.

Caghan Kizil *Technical University Dresden and DFG Center for Regenerative Therapies, Dresden, Germany.*
caghan.kizil@crt-dresden.de

Snapshot of a signalling complex

G-protein-coupled receptors initiate signalling pathways by forming complexes with agonist molecules and G proteins. The first crystal structure of such a complex is both reassuring and provocative. [SEE ARTICLE P.549](#) & [LETTER P.611](#)

THUE W. SCHWARTZ & THOMAS P. SAKMAR

The most widely used signal-transduction mechanism in nature involves G-protein-coupled receptors (GPCRs) embedded in cell membranes. Understanding precisely how this process occurs remains one of the most important and challenging questions in biology, not least because GPCRs act as sensors for a multitude of signals — from hormones and nutrients to pheromones and even light. And they're not just of interest to academics: GPCRs are the largest class of drug targets, with literally hundreds of discrete receptors amenable to pharmacological manipulation.

The binding of an agonist — a receptor activator — at the extracellular face of a GPCR causes the exchange of one nucleotide (GDP) for another (GTP) at a G protein bound to the receptor on the other side of the plasma membrane. The GTP-occupied α -subunit of the G protein and/or the $\beta\gamma$ -subunit then activates intracellular effector molecules. In a paper that will become an instant classic (page 549), Rasmussen *et al.*¹ present the long-awaited crystal structure of a basic transmembrane signalling complex: the agonist-bound β_2 -adrenergic receptor (β_2 AR) coupled to its G-protein partner, G_s . And in an accompanying paper on page 611 from the same group, Chung *et al.*² probe the protein dynamics of this signalling complex using a technique known as hydrogen-deuterium-exchange mass spectrometry (DXMS).

During the past two decades, a series of X-ray structures has provided great insight into the structural biology surrounding GPCR signalling. Of particular note are the recently reported structures showing the receptor part of the active signalling complex bound to small fragments of the G protein's α -subunit, or to antibodies that mimic the G protein^{3–5}. In addition, essential dynamic properties of the complex have been uncovered by spectroscopic and biophysical studies, many in the transduction system associated with vision^{6,7}. Importantly, however, Rasmussen *et al.*¹ are the first to have crystallized the complete complex of an active receptor and its G-protein partner.

There is a direct line connecting the fundamental, 30-year-old concept of 'allosteric'

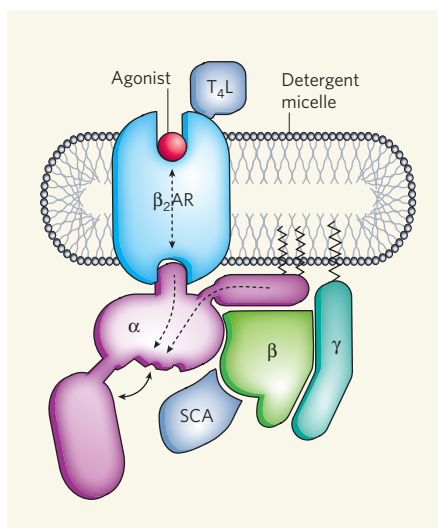


Figure 1 | Activated receptor-G-protein complex.

Transmembrane G-protein-coupled receptors (GPCRs) form complexes with agonist molecules and G proteins to initiate a wide range of signalling pathways in cells. Rasmussen *et al.*¹ report the first crystal structure of a complete, activated GPCR signalling complex: the agonist-bound β_2 -adrenergic receptor (β_2 AR) in complex with its G protein, G_s . In this schematic of the complex, α , β and γ are the three subunits of G_s . Also pictured are the protein T4-lysozyme (T_4L) and a single-chain antibody (SCA), which were incorporated into the complex to facilitate its crystallization. The detergent micelle acts in lieu of a cell membrane and maintains the structure of the complex. The zigzag lines on α and γ are endogenous lipid modifications that anchor the G protein to the micelle. Broken arrows within the α -subunit indicate structural changes that lead to the separation of its two domains (solid double-headed arrow), which allows the G protein to exchange one nucleotide (GDP) for another (GTP; nucleotides not shown). The broken, double-headed arrow in β_2 AR indicates the allosteric mechanism that connects G-protein binding to high-affinity agonist binding. (Graphic designed by Karina Åberg.)

coupling between receptor and G protein — and the corresponding effect of GDP on agonist binding^{8,9} — to Rasmussen and colleagues' crystallographic snapshot¹ of the GPCR-mediated signal-transduction process. Making the connection involved hard work rather than good luck — much of it old-fashioned, cold-room biochemistry based on

more than 20 years' worth of meticulously compiled tricks, skills and tools, not to mention considerable understanding of the β_2 AR system.

A complete review of the biochemical methods used to prepare the material that eventually yielded crystals for Rasmussen and co-workers' study would occupy an entire semester-long course in graduate school (and maybe it should). Selected highlights would include: the generation of a stable, nucleotide-free receptor-G-protein complex using an enzyme to remove GDP from the analogous GDP-containing complex; the development of antibodies from llamas that stabilize the active form of the G protein's α -subunit; and the use of single-particle electron microscopy to identify the most stable signalling complexes. These and similar approaches will certainly be useful in future work directed at improving the overall resolution of the structure of the agonist-bound β_2 AR- G_s complex — especially the receptor portion of the complex, which is currently visualized at relatively low resolution¹ — and for trapping other important forms of the complex, such as those involved in receptor deactivation.

The structure of Rasmussen and colleagues' complex (Fig. 1) and its mechanistic implications are striking, and would not all have been predicted. Surprisingly, the β - and γ -subunits of G_s barely make any direct contact with the receptor, but do facilitate interaction between the helical amino-terminal segment of the α -subunit of G_s and an intracellular loop (known as intracellular loop 2) of the receptor. The carboxy-terminal tail of the α -subunit engages the receptor in a cleft generated by the large outward movement of TM6, one of the receptor's seven transmembrane segments. The role of the α -subunit's C terminus was, in fact, clear from other work, such as the previously reported structures⁴ of rhodopsin (a GPCR involved in vision) in complex with a fragment of its corresponding G protein. However, the orientation and binding mode of this end of the α -subunit are different in the old⁴ and new¹ structures, showing that major conformational changes occur between the first encounter of the receptor with the GDP-occupied G protein and the formation of the nucleotide-free complex depicted by Rasmussen and colleagues' structure.

In contrast to Rasmussen *et al.*, who had to make several structural changes to the β_2 AR– G_s complex to facilitate its crystallization, Chung *et al.*² could perform their DXMS analysis on a mostly unmodified complex. Importantly, Chung and colleagues' study confirms most of the changes of conformation and interactions predicted from the X-ray structure¹. However, their results suggest that the RLLL peptide motif in G_s (a motif found in all G-protein α -subunits) undergoes considerable changes to its structure and/or stability during formation of the β_2 AR– G_s complex. This is at odds with the X-ray structure, in which the motif seems to be firmly locked in the middle of a β -sheet. Because the motif forms part of a protein strand (a β -strand) that directly connects intracellular loop 2 of the agonist-bound receptor with the 'P-loop' of G_s that binds to the β -phosphate in GDP, Chung and colleagues' data² provide a novel connection between receptor binding and nucleotide exchange that is not obvious from the X-ray structure¹.

The authors foresee that simultaneously, or possibly in rapid succession, the N-terminal helix of the α -subunit of G_s destabilizes the adjacent β -strand, and that the α -5 helix connected to the bound C-terminal tail becomes disordered. Both of these changes are communicated to the GDP-binding pocket of G_s , where two large domains of the protein's α -subunit open like a clamshell. This confirms a prediction made 18 years ago¹⁰, on the basis of the first X-ray structures of a G-protein α -subunit, that an activated receptor must open up the cleft in the α -subunit to allow nucleotide exchange. However, it seems that the 'clamshell' in Rasmussen and colleagues' structure may have been caught in an extreme, open form in their crystals, as a recent spectroscopic analysis⁷ of the rhodopsin–G-protein complex demonstrated that a more subtle, but still clear, opening occurs during nucleotide exchange.

For the first time, we have a detailed three-dimensional structure of the large (1,300 Å²) interface between an activated GPCR and its G-protein partner. We can now breathe a sigh of relief. But there will be no rest for weary workers in the GPCR field, because although the structure of the β_2 AR– G_s complex is reassuring, it is also provocative. Structural connections in the β_2 AR– G_s interface seem to be responsible for the allosteric mechanism that links receptor coupling to nucleotide exchange on the G protein. However, the structural basis for the increase in agonist affinity that occurs on G-protein binding is still elusive. Moreover, we still have no clear picture of what determines G-protein selectivity — the last five amino acids of a G-protein α -subunit determine its receptor binding partner¹¹, but Rasmussen and colleagues' β_2 AR– G_s structure¹ does not tell us how this happens.

Structures of additional receptor–G-protein complexes are therefore required, as are

complementary approaches of investigation, such as systems and molecular-dynamics simulations, and spectroscopic and other biophysical approaches^{6,7}. And we have not even mentioned GPCR kinases and β -arrestins, the proteins that are central to receptor desensitization and signalling through other intracellular pathways¹². But, for now, it's time to celebrate and enjoy a brief respite. Then it's back to the cold room. ■

Thue W. Schwartz is in the Laboratory for Molecular Pharmacology, Department of Neuroscience and Pharmacology, and at the NNF Center for Basic Metabolic Research, Faculty of Health Sciences, University of Copenhagen, DK-2200 Copenhagen, Denmark. **Thomas P. Sakmar** is in the Laboratory of Molecular Biology and Biochemistry,

The Rockefeller University, New York 10065, USA. e-mails: tws@sund.ku.dk; sakmar@rockefeller.edu

1. Rasmussen, S. G. F. *et al.* *Nature* **477**, 549–555 (2011).
2. Chung, K. Y. *et al.* *Nature* **477**, 611–615 (2011).
3. Rasmussen, S. G. F. *et al.* *Nature* **469**, 175–180 (2011).
4. Choe, H.-W. *et al.* *Nature* **471**, 651–655 (2011).
5. Standfuss, J. *et al.* *Nature* **471**, 656–660 (2011).
6. Huber, T. & Sakmar, T. P. *Trends Pharmacol. Sci.* **32**, 410–419 (2011).
7. Van Eps, N. *et al.* *Proc. Natl Acad. Sci. USA* **108**, 9420–9424 (2011).
8. Rodbell, M. *Nature* **284**, 17–22 (1980).
9. De Lean, A., Stadel, J. M. & Lefkowitz, R. J. *J. Biol. Chem.* **255**, 7108–7117 (1980).
10. Noel, J. P., Hamm, H. E. & Sigler, P. B. *Nature* **366**, 654–663 (1993).
11. Conklin, B. R., Farfel, Z., Lustig, K. D., Julius, D. & Bourne, H. R. *Nature* **363**, 274–276 (1993).
12. Lefkowitz, R. J. & Shenoy, S. K. *Science* **308**, 512–517 (2005).

COSMOLOGY

Gravity tested on large scales

Albert Einstein's general theory of relativity has passed a rigorous test on the scale of galaxy clusters. But the test does not rule out a popular alternative model of gravity. SEE LETTER P.567

GARY WEGNER

G gravity lies at the heart of understanding the cosmos. Tests of theories of gravity carried out in the laboratory — by British scientist Henry Cavendish, Hungarian physicist Roland von Eötvös, and a host of others up to the present day — are conceptually simple but are hampered by gravity's intrinsic weakness. Tests on the scale of our Solar System do better than laboratory tests, but clean tests of gravity on cosmic scales (those larger than individual galaxies) are difficult to perform. On page 567 of this issue, Wojtak *et al.*¹ describe how they have successfully tested Albert Einstein's theory of gravity, the general theory of relativity, against two well-known alternative theories of gravity on cosmic scales of the order of 1–10 megaparsecs. They did so by measuring the gravitational redshift of light from clusters of galaxies, an effect separate from the cosmological redshift, which is caused by the expansion of the Universe.

General relativity is the basis of our understanding of gravity. But twenty-first-century work in cosmology and particle theory strengthens the belief that it is an incomplete description. Although general relativity within the currently favoured 'Λ cold dark matter' (ΛCDM) cosmological model is undeniably highly successful in explaining cosmic-scale

observations (such as the cosmic microwave background radiation, baryonic oscillations, gravitational lensing and large-scale structure), many loose ends remain, ranging from assumptions about the nature of the Universe's dark matter and dark energy to details of galaxy formation and structure. Consequently, alternative theories of gravity have been constructed that attempt to eliminate the necessity for dark matter or dark energy in order to explain cosmic-scale observations.

Of the three 'classical' tests of general relativity — which involve measuring the bending of light by the Sun, the precession of the point of closest approach of Mercury to the Sun, and the gravitational redshift of light — the first two have been incorporated into a wider generation of new gravitational and cosmological measurements. Different gravity models can be compared with observations, with the result that general relativity is usually the winner. The third test (the gravitational redshift) measures a photon's frequency change as it moves in a gravitational potential. According to Clifford Will², it tests Einstein's equivalence principle rather than being a true test of general relativity, but implies that gravity is a 'metric' theory. Its terrestrial verifications have achieved high precision, but in the case of the Sun, measurements are dogged by pressure shifts and motions of the Sun's visible surface,

In contrast to Rasmussen *et al.*, who had to make several structural changes to the β_2 AR- G_s complex to facilitate its crystallization, Chung *et al.*² could perform their DXMS analysis on a mostly unmodified complex. Importantly, Chung and colleagues' study confirms most of the changes of conformation and interactions predicted from the X-ray structure¹. However, their results suggest that the RLLL peptide motif in G_s (a motif found in all G-protein α -subunits) undergoes considerable changes to its structure and/or stability during formation of the β_2 AR- G_s complex. This is at odds with the X-ray structure, in which the motif seems to be firmly locked in the middle of a β -sheet. Because the motif forms part of a protein strand (a β -strand) that directly connects intracellular loop 2 of the agonist-bound receptor with the 'P-loop' of G_s that binds to the β -phosphate in GDP, Chung and colleagues' data² provide a novel connection between receptor binding and nucleotide exchange that is not obvious from the X-ray structure¹.

The authors foresee that simultaneously, or possibly in rapid succession, the N-terminal helix of the α -subunit of G_s destabilizes the adjacent β -strand, and that the α -5 helix connected to the bound C-terminal tail becomes disordered. Both of these changes are communicated to the GDP-binding pocket of G_s , where two large domains of the protein's α -subunit open like a clamshell. This confirms a prediction made 18 years ago¹⁰, on the basis of the first X-ray structures of a G-protein α -subunit, that an activated receptor must open up the cleft in the α -subunit to allow nucleotide exchange. However, it seems that the 'clamshell' in Rasmussen and colleagues' structure may have been caught in an extreme, open form in their crystals, as a recent spectroscopic analysis⁷ of the rhodopsin-G-protein complex demonstrated that a more subtle, but still clear, opening occurs during nucleotide exchange.

For the first time, we have a detailed three-dimensional structure of the large (1,300 Å²) interface between an activated GPCR and its G-protein partner. We can now breathe a sigh of relief. But there will be no rest for weary workers in the GPCR field, because although the structure of the β_2 AR- G_s complex is reassuring, it is also provocative. Structural connections in the β_2 AR- G_s interface seem to be responsible for the allosteric mechanism that links receptor coupling to nucleotide exchange on the G protein. However, the structural basis for the increase in agonist affinity that occurs on G-protein binding is still elusive. Moreover, we still have no clear picture of what determines G-protein selectivity — the last five amino acids of a G-protein α -subunit determine its receptor binding partner¹¹, but Rasmussen and colleagues' β_2 AR- G_s structure¹ does not tell us how this happens.

Structures of additional receptor-G-protein complexes are therefore required, as are

complementary approaches of investigation, such as systems and molecular-dynamics simulations, and spectroscopic and other biophysical approaches^{6,7}. And we have not even mentioned GPCR kinases and β -arrestins, the proteins that are central to receptor desensitization and signalling through other intracellular pathways¹². But, for now, it's time to celebrate and enjoy a brief respite. Then it's back to the cold room. ■

Thue W. Schwartz is in the Laboratory for Molecular Pharmacology, Department of Neuroscience and Pharmacology, and at the NNF Center for Basic Metabolic Research, Faculty of Health Sciences, University of Copenhagen, DK-2200 Copenhagen, Denmark. **Thomas P. Sakmar** is in the Laboratory of Molecular Biology and Biochemistry,

The Rockefeller University, New York 10065, USA. e-mails: tws@sund.ku.dk; sakmar@rockefeller.edu

1. Rasmussen, S. G. F. *et al.* *Nature* **477**, 549–555 (2011).
2. Chung, K. Y. *et al.* *Nature* **477**, 611–615 (2011).
3. Rasmussen, S. G. F. *et al.* *Nature* **469**, 175–180 (2011).
4. Choe, H.-W. *et al.* *Nature* **471**, 651–655 (2011).
5. Standfuss, J. *et al.* *Nature* **471**, 656–660 (2011).
6. Huber, T. & Sakmar, T. P. *Trends Pharmacol. Sci.* **32**, 410–419 (2011).
7. Van Eps, N. *et al.* *Proc. Natl Acad. Sci. USA* **108**, 9420–9424 (2011).
8. Rodbell, M. *Nature* **284**, 17–22 (1980).
9. De Lean, A., Stadel, J. M. & Lefkowitz, R. J. *J. Biol. Chem.* **255**, 7108–7117 (1980).
10. Noel, J. P., Hamm, H. E. & Sigler, P. B. *Nature* **366**, 654–663 (1993).
11. Conklin, B. R., Farfel, Z., Lustig, K. D., Julius, D. & Bourne, H. R. *Nature* **363**, 274–276 (1993).
12. Lefkowitz, R. J. & Shenoy, S. K. *Science* **308**, 512–517 (2005).

COSMOLOGY

Gravity tested on large scales

Albert Einstein's general theory of relativity has passed a rigorous test on the scale of galaxy clusters. But the test does not rule out a popular alternative model of gravity. [SEE LETTER P.567](#)

GARY WEGNER

Gravity lies at the heart of understanding the cosmos. Tests of theories of gravity carried out in the laboratory — by British scientist Henry Cavendish, Hungarian physicist Roland von Eötvös, and a host of others up to the present day — are conceptually simple but are hampered by gravity's intrinsic weakness. Tests on the scale of our Solar System do better than laboratory tests, but clean tests of gravity on cosmic scales (those larger than individual galaxies) are difficult to perform. On page 567 of this issue, Wojtak *et al.*¹ describe how they have successfully tested Albert Einstein's theory of gravity, the general theory of relativity, against two well-known alternative theories of gravity on cosmic scales of the order of 1–10 megaparsecs. They did so by measuring the gravitational redshift of light from clusters of galaxies, an effect separate from the cosmological redshift, which is caused by the expansion of the Universe.

General relativity is the basis of our understanding of gravity. But twenty-first-century work in cosmology and particle theory strengthens the belief that it is an incomplete description. Although general relativity within the currently favoured 'Λ cold dark matter' (ΛCDM) cosmological model is undeniably highly successful in explaining cosmic-scale

observations (such as the cosmic microwave background radiation, baryonic oscillations, gravitational lensing and large-scale structure), many loose ends remain, ranging from assumptions about the nature of the Universe's dark matter and dark energy to details of galaxy formation and structure. Consequently, alternative theories of gravity have been constructed that attempt to eliminate the necessity for dark matter or dark energy in order to explain cosmic-scale observations.

Of the three 'classical' tests of general relativity — which involve measuring the bending of light by the Sun, the precession of the point of closest approach of Mercury to the Sun, and the gravitational redshift of light — the first two have been incorporated into a wider generation of new gravitational and cosmological measurements. Different gravity models can be compared with observations, with the result that general relativity is usually the winner. The third test (the gravitational redshift) measures a photon's frequency change as it moves in a gravitational potential. According to Clifford Will², it tests Einstein's equivalence principle rather than being a true test of general relativity, but implies that gravity is a 'metric' theory. Its terrestrial verifications have achieved high precision, but in the case of the Sun, measurements are dogged by pressure shifts and motions of the Sun's visible surface,

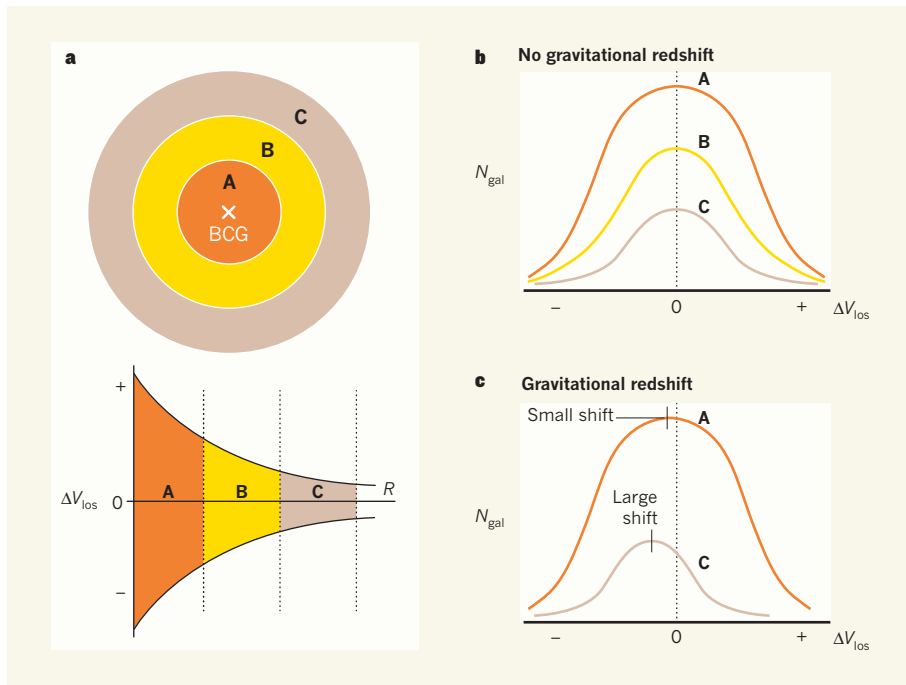


Figure 1 | Measuring the gravitational redshift of an idealized galaxy cluster. **a**, The cluster is centred on the brightest cluster galaxy (BCG) and sectioned into regions of increasing radius (here A–C). At corresponding radii (R), below, the velocities of cluster members relative to the velocity of the BCG, ΔV_{los} , lie within trumpet-like ‘caustic’ curves. **b**, Binning the velocities at the different radii gives velocity distributions; N_{gal} denotes the number of galaxies. Were there no gravitational redshift, the peaks of the distributions at different radii would average zero. **c**, With the gravitational redshift present, the peaks shift in velocity; the gravitational redshift is greatest at the cluster centre, so relative shifts become increasingly negative with radius. Wojtak *et al.*¹ constructed an averaged galaxy cluster based on archival data, measured its gravitational redshift, and compared the result with the predictions of general relativity and two other models of gravity.

the photosphere. Tests have also been carried out using white-dwarf stars, such as Sirius B and 40 Eri B, in binary star systems. But none of these tests probes cosmic distances.

Wojtak and colleagues¹ seem to be the first to successfully obtain gravitational redshifts for galaxy clusters. Although such gravitational redshifts have been discussed previously^{3,4}, there were insufficient galaxy data to produce significant results. For typical galaxy clusters, which are about 10^{14} times more massive than the Sun, the gravitational redshift corresponds to a velocity of the order of 10 kilometres per second. This is tiny in comparison with the spread of velocities of the constituent galaxies in a cluster. In some clusters the spread can reach $1,000 \text{ km s}^{-1}$, illustrating the difficulty of the measurement, to which other complications also contribute.

It is currently impracticable to determine the gravitational redshifts of individual clusters. Clusters have no ‘surface’, unlike the relatively simple case of the Sun, for which only a single measurement suffices, and the gravitational redshift varies with the line of sight through a cluster, as do the motions of the galaxies. This means that many galaxies per cluster are needed to determine the density and orbital structures of the cluster.

The term redshift can be confusing. The

measured redshift of an individual galaxy in a cluster is the combination of several variables. In addition to the gravitational redshift, there are the cosmological redshift, the effects of the galaxy’s motion in the cluster and the motion of the cluster itself. So, in order to determine the gravitational redshift, these other effects must be disentangled. Galaxy clusters are not always spherical, which adds to the complexity of interpreting the geometry of the galaxy velocities, and available data often provide only a handful of galaxy-velocity measurements per cluster. However, one might overcome this by combining galaxy clusters. Using simulations within the framework of the Λ CDM cosmological model, it was found⁵ that the averaging of about 2,500 clusters is required to compute a reliable gravitational redshift.

Wojtak *et al.* therefore constructed an averaged galaxy cluster. They used a subsample of a new galaxy cluster catalogue⁶ based on the Sloan Digital Sky Survey. They extracted 7,800 suitable galaxy clusters and 120,000 galaxy redshifts, which, as already noted, contain components of redshift other than the gravitational redshift. To extract these data, they defined a cluster member as lying within 6 megaparsecs of the cluster centre projected on the sky and within $\pm 4,000 \text{ km s}^{-1}$ in redshift, which statistically corresponds to a gravitationally bound

system. To determine each cluster’s centre, which is needed both to define the cluster members and to combine the clusters, they used the brightest cluster galaxy (BCG), a type of galaxy usually found near the cluster centre. Defining a cluster member’s redshift and velocity relative to the BCG allowed Wojtak *et al.* to eliminate the contributions of the cosmological redshift and cluster motion. This is an approximation, however. A cluster’s BCG can lie significantly far from the bottom of the cluster’s gravitational potential well and be moving, so the authors made a correction for the error produced by its not always being at the centre of the cluster.

After removing the contribution from contaminating background and foreground galaxies that are not associated with the clusters, Wojtak *et al.* assumed that the clusters are spherical on average, combined them and constructed the mean cluster, from which they could extract a density and a gravitational-redshift profile (Fig. 1). Given all these, albeit reasonable, assumptions, it is reassuring that the resulting gravitational redshifts in the averaged galaxy cluster are of the expected order of magnitude. Some readers might worry that the gravitational redshifts are smaller than the errors of measurement for each individual galaxy, and that small systematic errors and biases in the galaxy-velocity measurements could affect the result. As a check, Wojtak *et al.*¹ carried out Λ CDM numerical simulations on mock data and found that such effects did not significantly affect their results. Their gravitational redshifts are detected at the 99% confidence level.

Wojtak and colleagues¹ went on to compare the observed galaxy-cluster gravitational redshifts with the predictions of general relativity and modified gravity theories as a function of cluster radius. An important assumption needed for calculating the expected redshifts is that the cluster density follows the well-known Navarro–Frenk–White profile⁷, which the authors also found in their foregoing analysis from their measurements of the variation of the number of galaxies and the spread of their velocities with radius in the mean cluster.

The authors considered two popular alternative gravity models. One, known as $f(R)$ theory⁸, is consistent with general relativity on small scales and accounts for dark energy by invoking the ‘chameleon’ mechanism on cosmological scales; the other is the tensor–vector–scalar (TeVeS) theory⁹, which is a generalization of the modified Newtonian dynamics (MOND) model¹⁰ and is designed to eliminate the need for dark matter. Wojtak *et al.* found that general relativity fits best with the measured gravitational redshifts, that $f(R)$ theory is not ruled out, but that TeVeS/MOND is discrepant. Such a conclusion has been reached independently by other investigators, for example Reyes *et al.*¹¹, who analysed gravitational lensing and clustering obtained

from 70,000 galaxies in the Sloan Digital Sky Survey.

Discussions of gravity's properties will continue, but Wojtak and colleagues' pioneering work¹ gives a glimpse of the potential of new cosmological tests for achieving higher precision when millions of galaxy redshifts, from which gravitational redshifts can be extracted, become available in the future. ■

Gary Wegner is in the Department of Physics and Astronomy, Dartmouth College, Hanover, New Hampshire 03755, USA.
e-mail: gary.wegner@dartmouth.edu

1. Wojtak, R., Hansen, S. H. & Hjorth, J. *Nature* **477**, 567–569 (2011).
2. Will, C. M. *Living Rev. Relativity* **9**, 3 (2006).
3. Cappi, A. *Astron. Astrophys.* **301**, 6–10 (1995).
4. Broadhurst, T. & Scannapieco, E. *Astrophys. J.*

- 533**, L93–L97 (2000).
5. Kim, Y.-R. & Croft, R. A. C. *Astrophys. J.* **607**, 164–174 (2004).
6. Hao, J. et al. *Astrophys. J. Suppl. Ser.* **191**, 254–274 (2010).
7. Navarro, J. F., Frenk, C. S. & White, S. D. M. *Astrophys. J.* **490**, 493–508 (1997).
8. Skordis, C. *Class. Quantum Grav.* **26**, 143001 (2009).
9. De Felice, A. & Tsujikawa, S. *Living Rev. Relativity* **13**, 3 (2010).
10. Milgrom, M. *Astrophys. J.* **270**, 371–389 (1983).
11. Reyes, R. et al. *Nature* **464**, 256–258 (2010).

IMMUNOLOGY

Recognition of a unique partner

The mammalian immune system can fight a myriad of pathogens. In part, this involves a superfamily of cytoplasmic receptors that dictate assembly of specific pro-inflammatory inflammasome complexes. **SEE LETTERS P.592 & P.596**

DENISE M. MONACK

The innate immune system recognizes and responds to pathogens in a generic way. But it is not quite right to call it nonspecific, as is often done. Although certain aspects of innate immunity are non-specific, there is a growing appreciation that multicellular organisms have evolved specific recognition mechanisms as part of their innate immune defences to trigger clearance of pathogens. Two papers in this issue, by Kofoed and Vance¹ and by Zhao et al.², provide the latest evidence. They describe a family of intracellular receptors called NAIPs that ensure specific recognition of virulence-associated molecules — the molecules that are released by bacterial pathogens into the cytoplasm of the mouse immune cells in which they reside.

NAIPs belong to the NOD-like receptor (NLR) family of proteins, members of which have crucial roles in response to microbial infection. On detection of microbial components within the host-cell cytoplasm, NLRs initiate transcriptional programs that activate the immune response and trigger the activation of the protein-cleaving enzyme caspase-1. As part of a multiprotein complex called the inflammasome, caspase-1 cleaves and activates the pro-inflammatory immune mediators IL-1 β and IL-18, and initiates a pro-inflammatory cell-death process called pyroptosis³ (Fig. 1). Although previous genetic data³ support a crucial role for inflammasomes in immune defence and in inflammatory diseases, we are just beginning to understand the molecular mechanisms by which NLRs recognize specific foreign molecules and then induce the formation of inflammasomes.

Earlier studies^{3–5} showed that inflammasomes containing an NLR protein called

NLRC4 are activated in response to two distinct bacterial proteins. One is flagellin — which makes up the whip-like flagella that help bacteria to swim — and the other is PrgJ, an evolutionarily conserved component of bacterial type-III secretion systems, which mediates bacterial virulence. But how exactly the NLRC4 inflammasome responds to these different pathogen-associated proteins remained a mystery.

In a series of elegant experiments, Kofoed and Vance¹ (page 592) and Zhao and colleagues² (page 596) demonstrate that specific NAIPs bind directly to particular bacterial proteins. This leads to interaction between the NAIPs and NLRC4 and so to the activation of caspase-1. These results underscore the role of NAIPs in dictating the specificity of the NLRC4 inflammasome (Fig. 1).

The Vance group had previously shown⁴, in immune cells called macrophages, that NAIP5 is required for NLRC4-dependent activation of caspase-1 in response to *Legionella pneumophila* — a bacterial pathogen that causes Legionnaires' disease — and to the carboxy terminus of flagellin. But NAIP5 was not essential for caspase-1 activation in response to another flagellated intracellular bacterium, *Salmonella enterica* Typhimurium (which causes diarrhoea), or to its PrgJ protein⁴. As well as NAIP5, mouse macrophages express other NAIPs (NAIP1, NAIP2 and NAIP6), and so the authors proposed that each NAIP may have evolved to be specific for a unique bacterial ligand.

The new studies^{1,2} provide genetic evidence congruent with this proposal. Reducing the expression of the *Naip2* gene in isolated mouse macrophages prevented caspase-1 activation and pyroptosis in response to conserved rod proteins of type-III secretion

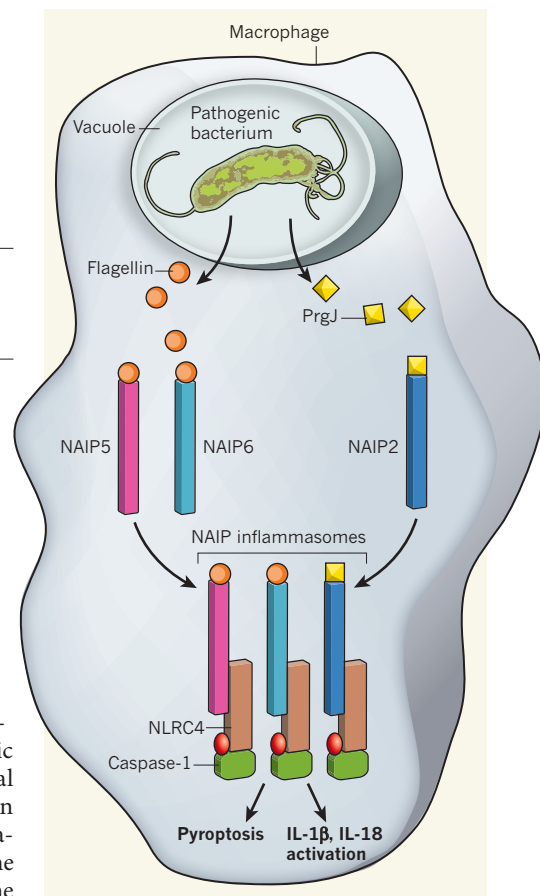


Figure 1 | NAIPs and ligand specificity of inflammasomes. Once engulfed by macrophages, pathogenic bacteria release virulence-associated molecules, such as PrgJ and flagellin, into the cytoplasm. New work^{1,2} in mouse cells shows that one of several related NAIP proteins, such as NAIP2, NAIP5 and NAIP6, binds specifically to such virulence-associated ligands. The NAIP–ligand complexes then bind directly to NLRC4, which is bound to the enzyme caspase-1 through a small adaptor molecule (red), thereby forming NAIP-containing inflammasomes. Caspase-1 is then activated, leading to activation of the pro-inflammatory immune mediators IL-1 β and IL-18 and to host-cell death by pyroptosis³.

systems of bacterial pathogens — PrgJ from *S. Typhimurium* and BsaK from *Burkholderia thailandensis*. This manipulation, however, did not affect pyroptosis in response to flagellin. Together with the authors' other findings⁴, the latest results indicate that NAIP2 is specifically required for activation of the NLRC4

from 70,000 galaxies in the Sloan Digital Sky Survey.

Discussions of gravity's properties will continue, but Wojtak and colleagues' pioneering work¹ gives a glimpse of the potential of new cosmological tests for achieving higher precision when millions of galaxy redshifts, from which gravitational redshifts can be extracted, become available in the future. ■

Gary Wegner is in the Department of Physics and Astronomy, Dartmouth College, Hanover, New Hampshire 03755, USA.
e-mail: gary.wegner@dartmouth.edu

1. Wojtak, R., Hansen, S. H. & Hjorth, J. *Nature* **477**, 567–569 (2011).
2. Will, C. M. *Living Rev. Relativity* **9**, 3 (2006).
3. Cappi, A. *Astron. Astrophys.* **301**, 6–10 (1995).
4. Broadhurst, T. & Scannapieco, E. *Astrophys. J.*

- 533**, L93–L97 (2000).
5. Kim, Y.-R. & Croft, R. A. C. *Astrophys. J.* **607**, 164–174 (2004).
6. Hao, J. et al. *Astrophys. J. Suppl. Ser.* **191**, 254–274 (2010).
7. Navarro, J. F., Frenk, C. S. & White, S. D. M. *Astrophys. J.* **490**, 493–508 (1997).
8. Skordis, C. *Class. Quantum Grav.* **26**, 143001 (2009).
9. De Felice, A. & Tsujikawa, S. *Living Rev. Relativity* **13**, 3 (2010).
10. Milgrom, M. *Astrophys. J.* **270**, 371–389 (1983).
11. Reyes, R. et al. *Nature* **464**, 256–258 (2010).

IMMUNOLOGY

Recognition of a unique partner

The mammalian immune system can fight a myriad of pathogens. In part, this involves a superfamily of cytoplasmic receptors that dictate assembly of specific pro-inflammatory inflammasome complexes. **SEE LETTERS P.592 & P.596**

DENISE M. MONACK

The innate immune system recognizes and responds to pathogens in a generic way. But it is not quite right to call it nonspecific, as is often done. Although certain aspects of innate immunity are non-specific, there is a growing appreciation that multicellular organisms have evolved specific recognition mechanisms as part of their innate immune defences to trigger clearance of pathogens. Two papers in this issue, by Kofoed and Vance¹ and by Zhao et al.², provide the latest evidence. They describe a family of intracellular receptors called NAIPs that ensure specific recognition of virulence-associated molecules — the molecules that are released by bacterial pathogens into the cytoplasm of the mouse immune cells in which they reside.

NAIPs belong to the NOD-like receptor (NLR) family of proteins, members of which have crucial roles in response to microbial infection. On detection of microbial components within the host-cell cytoplasm, NLRs initiate transcriptional programs that activate the immune response and trigger the activation of the protein-cleaving enzyme caspase-1. As part of a multiprotein complex called the inflammasome, caspase-1 cleaves and activates the pro-inflammatory immune mediators IL-1 β and IL-18, and initiates a pro-inflammatory cell-death process called pyroptosis³ (Fig. 1). Although previous genetic data³ support a crucial role for inflammasomes in immune defence and in inflammatory diseases, we are just beginning to understand the molecular mechanisms by which NLRs recognize specific foreign molecules and then induce the formation of inflammasomes.

Earlier studies^{3–5} showed that inflammasomes containing an NLR protein called

NLRC4 are activated in response to two distinct bacterial proteins. One is flagellin — which makes up the whip-like flagella that help bacteria to swim — and the other is PrgJ, an evolutionarily conserved component of bacterial type-III secretion systems, which mediates bacterial virulence. But how exactly the NLRC4 inflammasome responds to these different pathogen-associated proteins remained a mystery.

In a series of elegant experiments, Kofoed and Vance¹ (page 592) and Zhao and colleagues² (page 596) demonstrate that specific NAIPs bind directly to particular bacterial proteins. This leads to interaction between the NAIPs and NLRC4 and so to the activation of caspase-1. These results underscore the role of NAIPs in dictating the specificity of the NLRC4 inflammasome (Fig. 1).

The Vance group had previously shown⁴, in immune cells called macrophages, that NAIP5 is required for NLRC4-dependent activation of caspase-1 in response to *Legionella pneumophila* — a bacterial pathogen that causes Legionnaires' disease — and to the carboxy terminus of flagellin. But NAIP5 was not essential for caspase-1 activation in response to another flagellated intracellular bacterium, *Salmonella enterica* Typhimurium (which causes diarrhoea), or to its PrgJ protein⁴. As well as NAIP5, mouse macrophages express other NAIPs (NAIP1, NAIP2 and NAIP6), and so the authors proposed that each NAIP may have evolved to be specific for a unique bacterial ligand.

The new studies^{1,2} provide genetic evidence congruent with this proposal. Reducing the expression of the *Naip2* gene in isolated mouse macrophages prevented caspase-1 activation and pyroptosis in response to conserved rod proteins of type-III secretion

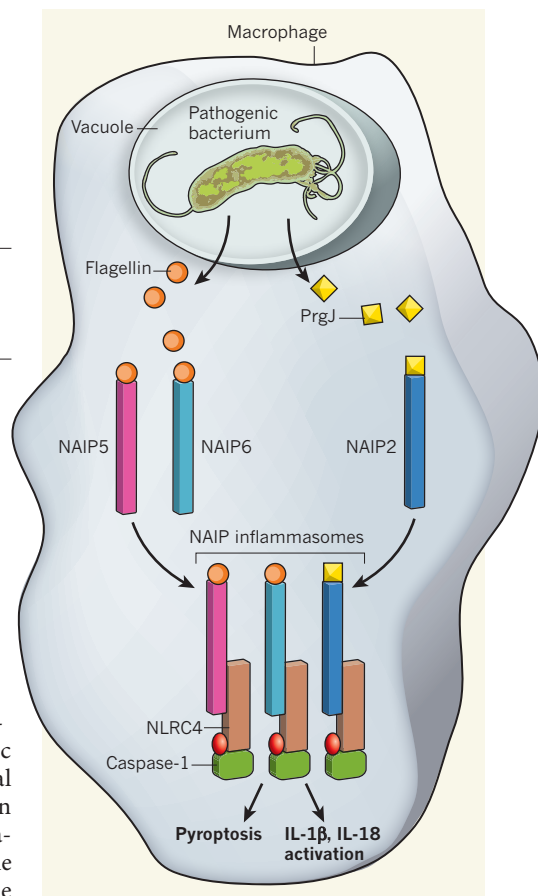


Figure 1 | NAIPs and ligand specificity of inflammasomes. Once engulfed by macrophages, pathogenic bacteria release virulence-associated molecules, such as PrgJ and flagellin, into the cytoplasm. New work^{1,2} in mouse cells shows that one of several related NAIP proteins, such as NAIP2, NAIP5 and NAIP6, binds specifically to such virulence-associated ligands. The NAIP–ligand complexes then bind directly to NLRC4, which is bound to the enzyme caspase-1 through a small adaptor molecule (red), thereby forming NAIP-containing inflammasomes. Caspase-1 is then activated, leading to activation of the pro-inflammatory immune mediators IL-1 β and IL-18 and to host-cell death by pyroptosis³.

systems of bacterial pathogens — PrgJ from *S. Typhimurium* and BsaK from *Burkholderia thailandensis*. This manipulation, however, did not affect pyroptosis in response to flagellin. Together with the authors' other findings⁴, the latest results indicate that NAIP2 is specifically required for activation of the NLRC4

inflammasome by the conserved components of the type-III secretion system, whereas NAIP5 and NAIP6 are specifically required for activation of the NLRC4 inflammasome by flagellin.

How inflammasomes are activated has been a matter of hot debate among immunologists. It was recently shown that, during infections, the intracellular receptor AIM2 binds to double-stranded viral or bacterial DNA in the host-cell cytoplasm, leading to the activation of an AIM2-containing inflammasome⁵. However, direct binding of ligands to NAIPs or to NLRC4 during bacterial infection had not been demonstrated.

Both teams^{1,2} now provide *in vitro* biochemical evidence that NAIP5 binds to flagellin and that NAIP2 binds to PrgJ/BsaK. Moreover, ligand binding led to the association of NAIP2 or NAIP5 with NLRC4, suggesting that NAIPs act before NLRC4 in the inflammatory pathway, dictating the specificity of inflammasome activation (Fig. 1). Although these findings shed light on some of the mechanisms

underlying inflammasome activation, it will be exciting to demonstrate ligand binding to NAIPs and the presence of NAIP-containing inflammasomes in infected macrophages.

As shown by the *in vitro* studies described above^{1,2}, one role of NAIPs is to recognize specific bacterial ligands in the cytoplasm of immune cells. But mice have multiple related *Naip* genes (*Naip1–7*), and a task for the future will be to determine the specific roles of their encoded proteins during infection. Studies in mice with altered *Naip* genes may determine whether individual NAIPs are expressed in a tissue- or cell-type-specific manner.

Unlike mice, humans have only one functional *NAIP* gene, which encodes the human NAIP protein. Zhao and colleagues² show that human NAIP does not respond to intracellular delivery of flagellin and BsaK/PrgJ-like rod proteins. Instead, it responds to CprI — the needle protein of the type-III secretion system from the bacterium *Chromobacterium violaceum*. In humans, *C. violaceum* infections

are rare. So further studies are required to determine the relevance of this NAIP specificity in humans. The genomic difference between humans and mice in terms of *Naip* copy number indicates divergent evolution of this particular arm of the innate immune system. Who knows, humans may have a superfamily of NLRs that determine inflammasome specificity in a similar way to the NAIPs in mice. ■

Denise M. Monack is in the Department of Microbiology and Immunology, Stanford School of Medicine, Stanford University, Stanford, California 94305, USA.
e-mail: dmonack@stanford.edu

1. Kofoed, E. M. & Vance, R. E. *Nature* **477**, 592–595 (2011).
2. Zhao, Y. *et al.* *Nature* **477**, 596–600 (2011).
3. Martinon, F., Mayor, A. & Tschopp, J. *Annu. Rev. Immunol.* **27**, 229–265 (2009).
4. Lightfield, K. L. *et al.* *Nature Immunol.* **9**, 1171–1178 (2008).
5. Broz, P. & Monack, D. M. *Immunol. Rev.* **243**, 174–190 (2011).

the ocean was motionless, this excluded the influence of ocean dynamics, thus disabling the Bjerknes feedback. Surprisingly, these models still reproduced the Southern Oscillation — the large-scale oscillation in sea-level atmospheric pressure that occurs between the eastern and western Pacific, and that is usually viewed as an atmospheric response to El Niño through the Bjerknes feedback.

The authors' models show that, in the absence of ocean dynamics, the state of the ocean's mixed layer — the uppermost layer that is affected by turbulence caused by wind, evaporative effects and so on — is driven by stochastic effects from the atmosphere. The ocean responds to these high-frequency signals from the atmosphere by turning them into lower-frequency effects, such as variations in SST. Statistical analysis of their data allowed the authors to plot SST spectra, which show how the magnitude of variation in SST changes with the period of the variation. In this way, the authors found that the variance of SST in their slab-ocean models increased with the period of the variation (to use the jargon, the SST spectrum was 'red').

The ocean's role in converting stochastic forcing into red noise signals was proposed some time ago by Klaus Hasselmann⁴, and has often been applied to interpret mid-latitude SST variations where the ocean is inert — that is, where the ocean has a thick mixed layer and slow subsurface-wave propagation — and where the atmosphere behaves randomly and is not sensitive to the SST. But Hasselmann's model has rarely been used to interpret atmospheric variability, and Clement *et al.* are the first to use it to explain the Southern Oscillation. The authors' results are reasonable and convincing.

CLIMATE SCIENCE

Ocean dynamics not required?

There are many challenges involved in understanding climate variations in the tropics, and how the Pacific climate might change with global warming. One study offers a surprising perspective on the mechanisms involved.

JING-JIA LUO

Tropical climatic phenomena, such as the anomalous surface warming of the central and eastern Pacific Ocean known as El Niño, often have significant environmental and socio-economic effects worldwide. Understanding their dynamics is vital to global climate prediction. Prevailing theories assume that the essential mechanism underpinning tropical climate is a dynamic ocean–atmosphere interaction¹. However, this idea is now being challenged, particularly for climatic variations that take place over more than just a few years. Writing in the *Journal of Climate*, Clement *et al.*² argue impressively that it is not necessary to couple ocean dynamics to the atmosphere in models to reproduce tropical climate modes and their associated global connections.

The atmosphere over the warm tropical Pacific is unstable. For instance, increases in sea surface temperature (SST) in the central and eastern regions can trigger excessive deep convection and large-scale adjustments to the circulation in the atmosphere, slackening the

easterly trade winds. The wind perturbations can drive ocean currents and propagate fast equatorial subsurface waves, which deepen the ocean's thermocline (a layer of water that separates the warm, upper part of the ocean from the cold, deeper water). This intensifies the initial increase in SST, particularly in the eastern Pacific, where the thermocline is shallow (about 50 metres below the surface), further weakening the trade winds. Thus, tropical climate signals such as increases in SST can grow rapidly to become some of the most important on Earth. The dynamic interactions between SST, trade winds and the thermocline are termed the Bjerknes feedback³, and form the foundation of our understanding of tropical-climate phenomena on seasonal and longer timescales.

But how essential are these interactions? Clement *et al.*² have investigated this issue by comparing the results from a pair of experiments, each of which simulated tropical climate using 13 models. In the first experiment, the models coupled the atmosphere thermodynamically to a motionless slab of ocean through heat and moisture fluxes. Because

inflammasome by the conserved components of the type-III secretion system, whereas NAIP5 and NAIP6 are specifically required for activation of the NLRC4 inflammasome by flagellin.

How inflammasomes are activated has been a matter of hot debate among immunologists. It was recently shown that, during infections, the intracellular receptor AIM2 binds to double-stranded viral or bacterial DNA in the host-cell cytoplasm, leading to the activation of an AIM2-containing inflammasome⁵. However, direct binding of ligands to NAIPs or to NLRC4 during bacterial infection had not been demonstrated.

Both teams^{1,2} now provide *in vitro* biochemical evidence that NAIP5 binds to flagellin and that NAIP2 binds to PrgJ/BsaK. Moreover, ligand binding led to the association of NAIP2 or NAIP5 with NLRC4, suggesting that NAIPs act before NLRC4 in the inflammatory pathway, dictating the specificity of inflammasome activation (Fig. 1). Although these findings shed light on some of the mechanisms

underlying inflammasome activation, it will be exciting to demonstrate ligand binding to NAIPs and the presence of NAIP-containing inflammasomes in infected macrophages.

As shown by the *in vitro* studies described above^{1,2}, one role of NAIPs is to recognize specific bacterial ligands in the cytoplasm of immune cells. But mice have multiple related *Naip* genes (*Naip1–7*), and a task for the future will be to determine the specific roles of their encoded proteins during infection. Studies in mice with altered *Naip* genes may determine whether individual NAIPs are expressed in a tissue- or cell-type-specific manner.

Unlike mice, humans have only one functional *NAIP* gene, which encodes the human NAIP protein. Zhao and colleagues² show that human NAIP does not respond to intracellular delivery of flagellin and BsaK/PrgJ-like rod proteins. Instead, it responds to CprI — the needle protein of the type-III secretion system from the bacterium *Chromobacterium violaceum*. In humans, *C. violaceum* infections

are rare. So further studies are required to determine the relevance of this NAIP specificity in humans. The genomic difference between humans and mice in terms of *Naip* copy number indicates divergent evolution of this particular arm of the innate immune system. Who knows, humans may have a superfamily of NLRs that determine inflammasome specificity in a similar way to the NAIPs in mice. ■

Denise M. Monack is in the Department of Microbiology and Immunology, Stanford School of Medicine, Stanford University, Stanford, California 94305, USA.
e-mail: dmonack@stanford.edu

1. Kofoed, E. M. & Vance, R. E. *Nature* **477**, 592–595 (2011).
2. Zhao, Y. *et al.* *Nature* **477**, 596–600 (2011).
3. Martinon, F., Mayor, A. & Tschopp, J. *Annu. Rev. Immunol.* **27**, 229–265 (2009).
4. Lightfield, K. L. *et al.* *Nature Immunol.* **9**, 1171–1178 (2008).
5. Broz, P. & Monack, D. M. *Immunol. Rev.* **243**, 174–190 (2011).

the ocean was motionless, this excluded the influence of ocean dynamics, thus disabling the Bjerknes feedback. Surprisingly, these models still reproduced the Southern Oscillation — the large-scale oscillation in sea-level atmospheric pressure that occurs between the eastern and western Pacific, and that is usually viewed as an atmospheric response to El Niño through the Bjerknes feedback.

The authors' models show that, in the absence of ocean dynamics, the state of the ocean's mixed layer — the uppermost layer that is affected by turbulence caused by wind, evaporative effects and so on — is driven by stochastic effects from the atmosphere. The ocean responds to these high-frequency signals from the atmosphere by turning them into lower-frequency effects, such as variations in SST. Statistical analysis of their data allowed the authors to plot SST spectra, which show how the magnitude of variation in SST changes with the period of the variation. In this way, the authors found that the variance of SST in their slab-ocean models increased with the period of the variation (to use the jargon, the SST spectrum was 'red').

The ocean's role in converting stochastic forcing into red noise signals was proposed some time ago by Klaus Hasselmann⁴, and has often been applied to interpret mid-latitude SST variations where the ocean is inert — that is, where the ocean has a thick mixed layer and slow subsurface-wave propagation — and where the atmosphere behaves randomly and is not sensitive to the SST. But Hasselmann's model has rarely been used to interpret atmospheric variability, and Clement *et al.* are the first to use it to explain the Southern Oscillation. The authors' results are reasonable and convincing.

CLIMATE SCIENCE

Ocean dynamics not required?

There are many challenges involved in understanding climate variations in the tropics, and how the Pacific climate might change with global warming. One study offers a surprising perspective on the mechanisms involved.

JING-JIA LUO

Tropical climatic phenomena, such as the anomalous surface warming of the central and eastern Pacific Ocean known as El Niño, often have significant environmental and socio-economic effects worldwide. Understanding their dynamics is vital to global climate prediction. Prevailing theories assume that the essential mechanism underpinning tropical climate is a dynamic ocean–atmosphere interaction¹. However, this idea is now being challenged, particularly for climatic variations that take place over more than just a few years. Writing in the *Journal of Climate*, Clement *et al.*² argue impressively that it is not necessary to couple ocean dynamics to the atmosphere in models to reproduce tropical climate modes and their associated global connections.

The atmosphere over the warm tropical Pacific is unstable. For instance, increases in sea surface temperature (SST) in the central and eastern regions can trigger excessive deep convection and large-scale adjustments to the circulation in the atmosphere, slackening the

easterly trade winds. The wind perturbations can drive ocean currents and propagate fast equatorial subsurface waves, which deepen the ocean's thermocline (a layer of water that separates the warm, upper part of the ocean from the cold, deeper water). This intensifies the initial increase in SST, particularly in the eastern Pacific, where the thermocline is shallow (about 50 metres below the surface), further weakening the trade winds. Thus, tropical climate signals such as increases in SST can grow rapidly to become some of the most important on Earth. The dynamic interactions between SST, trade winds and the thermocline are termed the Bjerknes feedback³, and form the foundation of our understanding of tropical-climate phenomena on seasonal and longer timescales.

But how essential are these interactions? Clement *et al.*² have investigated this issue by comparing the results from a pair of experiments, each of which simulated tropical climate using 13 models. In the first experiment, the models coupled the atmosphere thermodynamically to a motionless slab of ocean through heat and moisture fluxes. Because

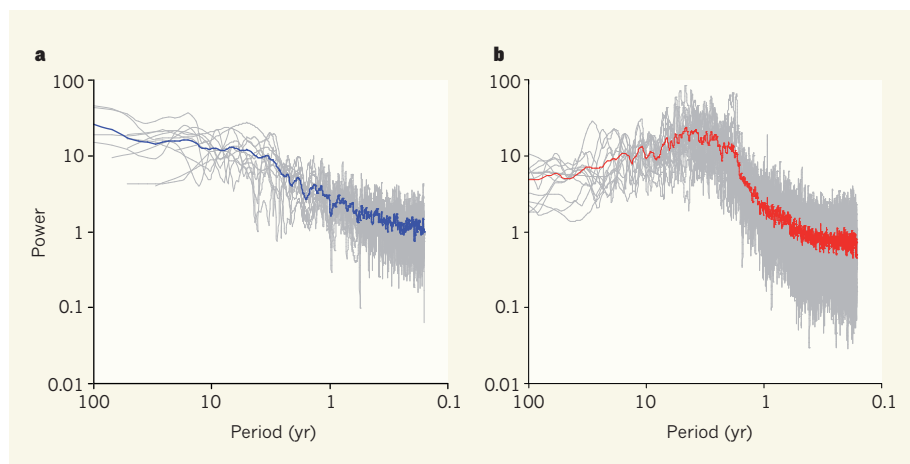


Figure 1 | Spectra of the Southern Oscillation. The Southern Oscillation is the large-scale oscillation in sea-level atmospheric pressure that occurs between the eastern and western Pacific, and is usually viewed as an atmospheric response to anomalous sea surface warming (El Niño) or cooling (La Niña) in the central and eastern Pacific. **a**, Clement *et al.*² have used 13 models in which the atmosphere is thermodynamically coupled to a motionless, 50-metre-deep slab of ocean to plot spectra of the Southern Oscillation (shown). The spectra plot power — a dimensionless measurement of the magnitude of the difference in atmospheric pressure between the western and eastern Pacific — against the period of the oscillation. The blue line denotes the multi-model average, whereas the grey lines indicate results from individual models. In these models, the magnitude of the Southern Oscillation increases with the period of the oscillation. **b**, The authors also plotted spectra using another 13 climate models in which full ocean dynamics are dynamically and thermodynamically coupled to the same atmospheric models as those used in **a**. These spectra show that the strongest variations in the Southern Oscillation occur roughly every 2–8 years. The difference between the two spectra indicates the role of coupled ocean dynamics in generating the variability of the Southern Oscillation at different timescales. (Data courtesy of A. Clement.)

Because the tropical atmosphere is influenced by SST, the ‘redness’ of SST variation is reflected in the spectra of the Southern Oscillation, which show how the magnitude of the difference in sea-level atmospheric pressure between the western and eastern tropical Pacific varies with the period of the oscillation (Fig. 1). Thus, interactions between the tropical atmosphere and ocean help to generate coupled climate modes. In the absence of the Bjerknes feedback, the authors’ models describe a robust, thermally coupled mode whose structure resembles the El Niño–Southern Oscillation (ENSO) on both interannual timescales (periods in the range of two to ten years) and decadal or multidecadal timescales.

Do Clement and colleagues’ findings suggest that the Bjerknes feedback and ocean dynamics are not important for tropical climate? Not necessarily. In their second experiment, they simulated tropical climate using 13 models in which the atmosphere is coupled to a fully dynamic ocean². The interannual ENSO signal predicted by these models was, on average, two to three times stronger than that predicted by the slab-ocean models. What’s more, a previous study⁵ found that 19 out of 24 computer models lacking coupled ocean dynamics severely underestimate the intensity of ENSO. These results, taken together with Clement and colleagues’ findings², explicitly affirm the classic ENSO theory involving coupled ocean dynamics that underpins extensive current research into ENSO and global-climate forecasts on

seasonal-to-interannual timescales. However, on decadal and longer timescales, ocean dynamics seem to be less important than was thought: the magnitude of the variance in SST in the climate mode observed by Clement *et al.*² in their ocean-slab models is comparable to that observed in reality. When the authors coupled full ocean dynamics to the atmosphere in their models, this actually weakened the decadal and multidecadal variability by constraining the variance of tropical climate to interannual timescales.

The underlying mechanisms of decadal and multidecadal climate variation in the tropical Pacific are unknown. A lack of ocean observations means that the precise roles of the ocean — particularly the role of subsurface signals originating from the extratropics, the regions poleward of the tropics — remain controversial. There is even debate over whether the decadal ENSO-like mode is a physical entity or an artefact arising from the nonlinearity of ocean–atmosphere coupling⁶. Similar debates have been sparked by attempts to detect and project tropical Pacific climate change under global warming^{7,8}: will increased anthropogenic greenhouse-gas emissions lead to an El Niño-like state, or the opposite (a La Niña-like state)? Clement and colleagues’ work supports neither of the two scenarios, but questions the importance of the Bjerknes feedback, the core element of prevailing hypotheses.

One limitation of the authors’ findings² is that many of their climate models are severely flawed when simulating the period and



50 Years Ago

The circumstances of our age seem to present us with a very real problem which may fittingly be discussed under the title of “Science and the Classics”. In response to an international challenge we are engaged in expanding our educational services. The urgent need is for more scientists, more technologists, more engineers. This means a relative, if not an absolute, decline in the study of the humanities. Is this shift in the character of our education good or bad? Is there anything in the claim that the study of the humanities produced a more balanced type of man? Is there in the scientific mode of training any tendency against which it is desirable to be on our guard? I think that there is a tendency among scientists to suppose that the physical world alone is real, and that the very successful methods they have devised for dealing with it are applicable to every aspect of life. This, to my way of thinking, is a dangerous illusion.

From *Nature* 30 September 1961

100 Years Ago

Exhaustive tests have been made during the last two weeks by Mr. A. W. Sharman with instruments invented by him for telephoning through water without wires. A small telephone station has been erected in a room in an hotel on the cliffs at Pegwell Bay, and the other station has been fitted up on a motor-boat cruising in various parts of the bay ... The speech transmitted through the water has been very distinct, and the system has shown good possibilities of its being used as a means of verbal communication between two ships, such as a battleship and a submarine.

From *Nature* 28 September 1911

magnitude of ENSO. The misrepresentation of ENSO dynamics in such models may preclude an accurate separation of dynamic and thermodynamic coupling effects. Moreover, most climate models erroneously predict the existence of an intertropical convergence zone (ITCZ) — a band near the Equator where the trade winds from the two hemispheres converge — in the South Pacific, in addition to the real one observed north of the Equator. This problem is known as double ITCZ bias. By raising the oceanic thermocline, ITCZs can effectively prevent extratropical ocean signals from penetrating into the equatorial Pacific⁹, both in models and in the real world. Therefore,

the ocean's role in decadal and multidecadal ENSO-like variations might be underestimated in models featuring double ITCZ bias. Increasing the space and time coverage of ocean observations will be crucial to address this possibility. Similar experiments to those of Clement *et al.*, conducted using improved climate models, are also needed to validate and refine the current results². Despite these caveats, the authors' findings may help to reshape our view not only of today's climate, but also of past and future changes in tropical climate. ■

Jing-Jia Luo is at the Research Institute for Global Change, JAMSTEC, 3173-25

Showa-machi, Kanazawa-ku, Yokohama, Kanagawa 236-0001, Japan.
e-mail: luo@jamstec.go.jp

1. Neelin, J. D. *et al.* *J. Geophys. Res.* <http://dx.doi.org/10.1029/97JC03424> (1998).
2. Clement, A., DiNezio, P. & Deser, C. *J. Clim.* **24**, 4056–4072 (2011).
3. Bjerknes, J. *Mon. Wea. Rev.* **97**, 163–172 (1969).
4. Hasselmann, K. *Tellus* **28**, 473–485 (1976).
5. Dommenget, D. *Geophys. Res. Lett.* <http://dx.doi.org/10.1029/2010GL044888> (2010).
6. Timmermann, A. & Jin, F.-F. *Geophys. Res. Lett.* <http://dx.doi.org/10.1029/2001GL013369> (2002).
7. Vecchi, G. A. & Soden, B. J. *J. Clim.* **20**, 4316–4340 (2007).
8. Karnauskas, K. B. *et al.* *J. Clim.* **22**, 4316–4321 (2009).
9. Luo, J.-J. & Yamagata, T. *J. Geophys. Res.* <http://dx.doi.org/10.1029/2000JC000471> (2001).

general, SUMOylation does not mark substrate proteins for degradation, whereas ubiquitination does. The two tags may even counteract one another by competing for the same lysine residues to define the fate of substrate proteins⁶. Discovered in 1996, some 16 years after ubiquitination, SUMOylation has quickly proved its importance in processes as diverse as protein transport, gene transcription, DNA replication and repair, and chromosome segregation^{6–8}.

Both SUMO and ubiquitin are activated by an enzyme called E1 to associate with a carrier/conjugating enzyme, E2; they are both appended to their substrates by a ligating enzyme, E3. The striking difference in the two pathways is that, whereas SUMO conjugation uses a single E2, roughly 40 E2s are active in ubiquitination. Furthermore, only one isoform of ubiquitin exists in cells, but there are four SUMO isoforms — of which SUMOs 1–3 could be involved in substrate SUMOylation⁶.

Kho *et al.*¹ have discovered that SERCA2a in the heart is SUMOylated by SUMO1 at two lysine residues (at positions 480 and 585) within its ATP-binding domain¹. They found that mutating both lysines prevents SUMOylation, enhances ubiquitination and degradation of SERCA2a and — surprisingly — reduces ATP binding. They conclude that SUMOylation of SERCA2a could be a prerequisite for the protein's stability as well as for its Ca²⁺-pumping activity.

The authors found that failing hearts have decreased amounts of both SUMO1 and SERCA2a. In an animal model of heart failure, overexpression of SERCA2a or SUMO1 led to a reversal of heart failure; however, in an animal model deficient in SERCA2a, re-expression of SUMO1 alone did not restore heart function. Although several cardiac proteins (transcription factors, for example) and processes (such as gene transcription) could be affected by reduced levels of SUMO1, Kho and colleagues' findings indicate that only the recovery of SERCA2a SUMOylation can reverse deterioration of the heart. Another cardiac protein, lamin A, also undergoes SUMOylation at sites

CARDIOVASCULAR BIOLOGY

Heart fails without pump partner

Mishandling of calcium ions by cardiac cells causes the heart to malfunction. The discovery of a crucial modification to a calcium pump inside the cell opens up a potential way to correct this. SEE LETTER P.601

SUDHA K. SHENOY & HOWARD A. ROCKMAN

The activity of SERCA2a, a crucial membrane protein located inside cardiac muscle cells (myocytes), is compromised in heart failure. SERCA2a acts as a calcium-ion pump, being responsible for the uptake of calcium ions during excitation–contraction coupling, a process that connects electrical excitation of myocytes with contraction of the heart muscle. On page 601 of this issue¹, Kho and colleagues show that SERCA2a, which is present on the membrane of a cellular organelle known as the sarcoplasmic reticulum, is modified after its translation by the attachment of another protein known as SUMO1. This 'SUMOylation' stabilizes the SERCA2a pump and maintains normal functioning of the heart, which is disrupted if the attachment fails.

Faulty flux of calcium ions (Ca²⁺) inside the cardiac myocyte is a central cause of contractile dysfunction and arrhythmias in pathophysiological heart conditions. This is because Ca²⁺ ions are essential for cardiac electrical activity and directly activate the elements of the myocyte that cause contraction.

Figure 1 shows how these events occur in a normal cardiac myocyte. In response to the wave of depolarization generated by the arrival of an action potential at the myocyte membrane, Ca²⁺ ions enter the cell through voltage-gated L-type Ca²⁺ channels in the membrane. This 'activator' Ca²⁺ triggers the release of a large amount of Ca²⁺ stored in the

sarcoplasmic reticulum in a process known as Ca²⁺-induced Ca²⁺ release, as well as generating a small Ca²⁺-release event from the sarcoplasmic reticulum (a Ca²⁺ 'spark')^{2,3}. During excitation–contraction coupling, several thousand Ca²⁺ sparks in each cell are synchronized by the action potential to increase intracellular Ca²⁺, which bathes the muscle's contractile myofilaments and causes the heart cell to contract. To initiate relaxation of the heart, Ca²⁺ is removed from the cytoplasm, primarily through the action of SERCA2a (ref. 2), which hydrolyses the cellular energy molecule ATP to pump Ca²⁺ ions back into the sarcoplasmic reticulum.

SERCA2a activity is stimulated by the hormone noradrenaline, released from sympathetic nerves, through phosphorylation of a protein called phospholamban^{2,4} and by the SUMOylation mechanism reported by Kho *et al.*¹ (Fig. 1). Deficits in SERCA2a activity and expression are associated with heart failure in animal models and in humans, indicating that it could be a critical molecule in the pathogenesis of cardiac failure⁵.

SUMO1 (small ubiquitin-like modifier 1) is related to the well-known protein tag ubiquitin. Their three-dimensional structures are remarkably similar, despite there being little sequence conservation between them⁶. Both SUMOylation and ubiquitination involve dynamic and reversible covalent post-translational modification of lysine amino-acid residues in their substrate proteins, leading to changes in stability, activity and/or subcellular localization of the modified proteins. In

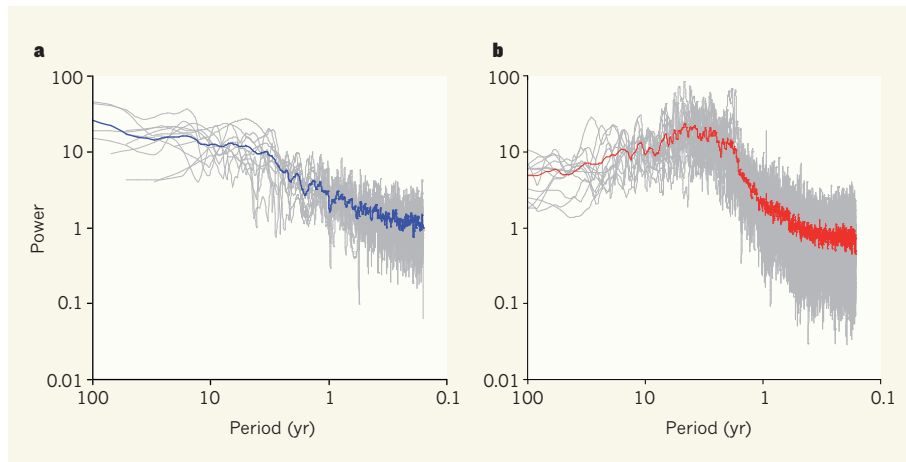


Figure 1 | Spectra of the Southern Oscillation. The Southern Oscillation is the large-scale oscillation in sea-level atmospheric pressure that occurs between the eastern and western Pacific, and is usually viewed as an atmospheric response to anomalous sea surface warming (El Niño) or cooling (La Niña) in the central and eastern Pacific. **a**, Clement *et al.*² have used 13 models in which the atmosphere is thermodynamically coupled to a motionless, 50-metre-deep slab of ocean to plot spectra of the Southern Oscillation (shown). The spectra plot power — a dimensionless measurement of the magnitude of the difference in atmospheric pressure between the western and eastern Pacific — against the period of the oscillation. The blue line denotes the multi-model average, whereas the grey lines indicate results from individual models. In these models, the magnitude of the Southern Oscillation increases with the period of the oscillation. **b**, The authors also plotted spectra using another 13 climate models in which full ocean dynamics are dynamically and thermodynamically coupled to the same atmospheric models as those used in **a**. These spectra show that the strongest variations in the Southern Oscillation occur roughly every 2–8 years. The difference between the two spectra indicates the role of coupled ocean dynamics in generating the variability of the Southern Oscillation at different timescales. (Data courtesy of A. Clement.)

Because the tropical atmosphere is influenced by SST, the ‘redness’ of SST variation is reflected in the spectra of the Southern Oscillation, which show how the magnitude of the difference in sea-level atmospheric pressure between the western and eastern tropical Pacific varies with the period of the oscillation (Fig. 1). Thus, interactions between the tropical atmosphere and ocean help to generate coupled climate modes. In the absence of the Bjerknes feedback, the authors’ models describe a robust, thermally coupled mode whose structure resembles the El Niño–Southern Oscillation (ENSO) on both interannual timescales (periods in the range of two to ten years) and decadal or multidecadal timescales.

Do Clement and colleagues’ findings suggest that the Bjerknes feedback and ocean dynamics are not important for tropical climate? Not necessarily. In their second experiment, they simulated tropical climate using 13 models in which the atmosphere is coupled to a fully dynamic ocean². The interannual ENSO signal predicted by these models was, on average, two to three times stronger than that predicted by the slab-ocean models. What’s more, a previous study⁵ found that 19 out of 24 computer models lacking coupled ocean dynamics severely underestimate the intensity of ENSO. These results, taken together with Clement and colleagues’ findings², explicitly affirm the classic ENSO theory involving coupled ocean dynamics that underpins extensive current research into ENSO and global-climate forecasts on

seasonal-to-interannual timescales. However, on decadal and longer timescales, ocean dynamics seem to be less important than was thought: the magnitude of the variance in SST in the climate mode observed by Clement *et al.*² in their ocean-slab models is comparable to that observed in reality. When the authors coupled full ocean dynamics to the atmosphere in their models, this actually weakened the decadal and multidecadal variability by constraining the variance of tropical climate to interannual timescales.

The underlying mechanisms of decadal and multidecadal climate variation in the tropical Pacific are unknown. A lack of ocean observations means that the precise roles of the ocean — particularly the role of subsurface signals originating from the extratropics, the regions poleward of the tropics — remain controversial. There is even debate over whether the decadal ENSO-like mode is a physical entity or an artefact arising from the nonlinearity of ocean–atmosphere coupling⁶. Similar debates have been sparked by attempts to detect and project tropical Pacific climate change under global warming^{7,8}: will increased anthropogenic greenhouse-gas emissions lead to an El Niño-like state, or the opposite (a La Niña-like state)? Clement and colleagues’ work supports neither of the two scenarios, but questions the importance of the Bjerknes feedback, the core element of prevailing hypotheses.

One limitation of the authors’ findings² is that many of their climate models are severely flawed when simulating the period and



50 Years Ago

The circumstances of our age seem to present us with a very real problem which may fittingly be discussed under the title of “Science and the Classics”. In response to an international challenge we are engaged in expanding our educational services. The urgent need is for more scientists, more technologists, more engineers. This means a relative, if not an absolute, decline in the study of the humanities. Is this shift in the character of our education good or bad? Is there anything in the claim that the study of the humanities produced a more balanced type of man? Is there in the scientific mode of training any tendency against which it is desirable to be on our guard? I think that there is a tendency among scientists to suppose that the physical world alone is real, and that the very successful methods they have devised for dealing with it are applicable to every aspect of life. This, to my way of thinking, is a dangerous illusion.

From *Nature* 30 September 1961

100 Years Ago

Exhaustive tests have been made during the last two weeks by Mr. A. W. Sharman with instruments invented by him for telephoning through water without wires. A small telephone station has been erected in a room in an hotel on the cliffs at Pegwell Bay, and the other station has been fitted up on a motor-boat cruising in various parts of the bay ... The speech transmitted through the water has been very distinct, and the system has shown good possibilities of its being used as a means of verbal communication between two ships, such as a battleship and a submarine.

From *Nature* 28 September 1911

magnitude of ENSO. The misrepresentation of ENSO dynamics in such models may preclude an accurate separation of dynamic and thermodynamic coupling effects. Moreover, most climate models erroneously predict the existence of an intertropical convergence zone (ITCZ) — a band near the Equator where the trade winds from the two hemispheres converge — in the South Pacific, in addition to the real one observed north of the Equator. This problem is known as double ITCZ bias. By raising the oceanic thermocline, ITCZs can effectively prevent extratropical ocean signals from penetrating into the equatorial Pacific⁹, both in models and in the real world. Therefore,

the ocean's role in decadal and multidecadal ENSO-like variations might be underestimated in models featuring double ITCZ bias. Increasing the space and time coverage of ocean observations will be crucial to address this possibility. Similar experiments to those of Clement *et al.*, conducted using improved climate models, are also needed to validate and refine the current results². Despite these caveats, the authors' findings may help to reshape our view not only of today's climate, but also of past and future changes in tropical climate. ■

Jing-Jia Luo is at the Research Institute for Global Change, JAMSTEC, 3173-25

Showa-machi, Kanazawa-ku, Yokohama, Kanagawa 236-0001, Japan.
e-mail: luo@jamstec.go.jp

1. Neelin, J. D. *et al.* *J. Geophys. Res.* <http://dx.doi.org/10.1029/97JC03424> (1998).
2. Clement, A., DiNezio, P. & Deser, C. *J. Clim.* **24**, 4056–4072 (2011).
3. Bjerknes, J. *Mon. Wea. Rev.* **97**, 163–172 (1969).
4. Hasselmann, K. *Tellus* **28**, 473–485 (1976).
5. Dommenget, D. *Geophys. Res. Lett.* <http://dx.doi.org/10.1029/2010GL044888> (2010).
6. Timmermann, A. & Jin, F.-F. *Geophys. Res. Lett.* <http://dx.doi.org/10.1029/2001GL013369> (2002).
7. Vecchi, G. A. & Soden, B. J. *J. Clim.* **20**, 4316–4340 (2007).
8. Karnauskas, K. B. *et al.* *J. Clim.* **22**, 4316–4321 (2009).
9. Luo, J.-J. & Yamagata, T. *J. Geophys. Res.* <http://dx.doi.org/10.1029/2000JC000471> (2001).

general, SUMOylation does not mark substrate proteins for degradation, whereas ubiquitination does. The two tags may even counteract one another by competing for the same lysine residues to define the fate of substrate proteins⁶. Discovered in 1996, some 16 years after ubiquitination, SUMOylation has quickly proved its importance in processes as diverse as protein transport, gene transcription, DNA replication and repair, and chromosome segregation^{6–8}.

Both SUMO and ubiquitin are activated by an enzyme called E1 to associate with a carrier/conjugating enzyme, E2; they are both appended to their substrates by a ligating enzyme, E3. The striking difference in the two pathways is that, whereas SUMO conjugation uses a single E2, roughly 40 E2s are active in ubiquitination. Furthermore, only one isoform of ubiquitin exists in cells, but there are four SUMO isoforms — of which SUMOs 1–3 could be involved in substrate SUMOylation⁶.

Kho *et al.*¹ have discovered that SERCA2a in the heart is SUMOylated by SUMO1 at two lysine residues (at positions 480 and 585) within its ATP-binding domain¹. They found that mutating both lysines prevents SUMOylation, enhances ubiquitination and degradation of SERCA2a and — surprisingly — reduces ATP binding. They conclude that SUMOylation of SERCA2a could be a prerequisite for the protein's stability as well as for its Ca²⁺-pumping activity.

The authors found that failing hearts have decreased amounts of both SUMO1 and SERCA2a. In an animal model of heart failure, overexpression of SERCA2a or SUMO1 led to a reversal of heart failure; however, in an animal model deficient in SERCA2a, re-expression of SUMO1 alone did not restore heart function. Although several cardiac proteins (transcription factors, for example) and processes (such as gene transcription) could be affected by reduced levels of SUMO1, Kho and colleagues' findings indicate that only the recovery of SERCA2a SUMOylation can reverse deterioration of the heart. Another cardiac protein, lamin A, also undergoes SUMOylation at sites

CARDIOVASCULAR BIOLOGY

Heart fails without pump partner

Mishandling of calcium ions by cardiac cells causes the heart to malfunction. The discovery of a crucial modification to a calcium pump inside the cell opens up a potential way to correct this. SEE LETTER P.601

SUDHA K. SHENOY & HOWARD A. ROCKMAN

The activity of SERCA2a, a crucial membrane protein located inside cardiac muscle cells (myocytes), is compromised in heart failure. SERCA2a acts as a calcium-ion pump, being responsible for the uptake of calcium ions during excitation–contraction coupling, a process that connects electrical excitation of myocytes with contraction of the heart muscle. On page 601 of this issue¹, Kho and colleagues show that SERCA2a, which is present on the membrane of a cellular organelle known as the sarcoplasmic reticulum, is modified after its translation by the attachment of another protein known as SUMO1. This 'SUMOylation' stabilizes the SERCA2a pump and maintains normal functioning of the heart, which is disrupted if the attachment fails.

Faulty flux of calcium ions (Ca²⁺) inside the cardiac myocyte is a central cause of contractile dysfunction and arrhythmias in pathophysiological heart conditions. This is because Ca²⁺ ions are essential for cardiac electrical activity and directly activate the elements of the myocyte that cause contraction.

Figure 1 shows how these events occur in a normal cardiac myocyte. In response to the wave of depolarization generated by the arrival of an action potential at the myocyte membrane, Ca²⁺ ions enter the cell through voltage-gated L-type Ca²⁺ channels in the membrane. This 'activator' Ca²⁺ triggers the release of a large amount of Ca²⁺ stored in the

sarcoplasmic reticulum in a process known as Ca²⁺-induced Ca²⁺ release, as well as generating a small Ca²⁺-release event from the sarcoplasmic reticulum (a Ca²⁺ 'spark')^{2,3}. During excitation–contraction coupling, several thousand Ca²⁺ sparks in each cell are synchronized by the action potential to increase intracellular Ca²⁺, which bathes the muscle's contractile myofilaments and causes the heart cell to contract. To initiate relaxation of the heart, Ca²⁺ is removed from the cytoplasm, primarily through the action of SERCA2a (ref. 2), which hydrolyses the cellular energy molecule ATP to pump Ca²⁺ ions back into the sarcoplasmic reticulum.

SERCA2a activity is stimulated by the hormone noradrenaline, released from sympathetic nerves, through phosphorylation of a protein called phospholamban^{2,4} and by the SUMOylation mechanism reported by Kho *et al.*¹ (Fig. 1). Deficits in SERCA2a activity and expression are associated with heart failure in animal models and in humans, indicating that it could be a critical molecule in the pathogenesis of cardiac failure⁵.

SUMO1 (small ubiquitin-like modifier 1) is related to the well-known protein tag ubiquitin. Their three-dimensional structures are remarkably similar, despite there being little sequence conservation between them⁶. Both SUMOylation and ubiquitination involve dynamic and reversible covalent post-translational modification of lysine amino-acid residues in their substrate proteins, leading to changes in stability, activity and/or subcellular localization of the modified proteins. In

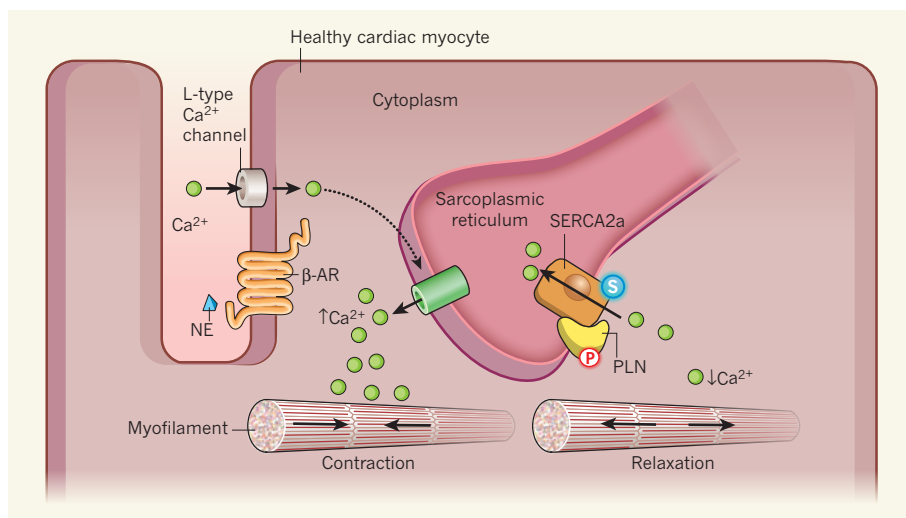


Figure 1 | The SERCA2a calcium pump and its modification by SUMO1. In a healthy cardiac myocyte, noradrenaline (NE) released from sympathetic nerves stimulates β -adrenergic receptors (β -AR) to enhance the activity of L-type calcium-ion (Ca^{2+}) channels on the cell membrane, leading to Ca^{2+} entry into the cell. This in turn triggers the release of a large amount of Ca^{2+} from the sarcoplasmic reticulum into the cytoplasm, and leads to myofilament contraction. Noradrenaline stimulation also causes phosphorylation (P) of phospholamban (PLN), releasing its inhibitory effect on the sarcoplasmic-reticulum Ca^{2+} -ATPase pump, SERCA2a. Activated SERCA2a binds Ca^{2+} ions from the cytoplasm and pumps them back into the sarcoplasmic reticulum, allowing the myofilaments to relax. Kho *et al.*¹ show that only SERCA2a modified by SUMO1 (S) can transport Ca^{2+} ions and maintain a healthy heart. In a failing heart (not shown), SUMO1 is depleted and SERCA2a remains in the un-SUMOylated form. It therefore cannot bind ATP and cannot effectively pump Ca^{2+} back into the sarcoplasmic-reticulum lumen, so relaxation of cardiac muscle is impaired.

that are known to be mutated in patients with familial dilated cardiomyopathy and those with conduction-system disease, suggesting that these disorders may also be attributable to defective SUMOylation⁹.

However, SUMO1 deficiency also causes degradation of the transcription factor SRF by caspase enzymes⁹, which are involved in programmed cell death. Therefore, restoring SUMO1 activity by introducing the SUMO1 transgene may block this and other cell-death pathways in the failing heart and lead to the observed recovery. Further investigation is needed to determine the cause of the reduced SUMO1 levels seen in heart failure: the protein does not undergo degradation and there is no evidence that gene transcription is inhibited.

In contrast to the stimulation of SERCA2a by SUMO1, some potassium-ion channels in the cell membrane are inhibited by SUMOylation^{10,11}; they regain their ion-selective activity when SUMOylation is blocked or reversed. Furthermore, there is evidence that SUMOylation could be critical for transport processes into the nucleus^{6,12}. It might therefore also facilitate Ca^{2+} transport across membranes, for example by imparting structural features of an 'on/off' switch to the ion channels.

Kho *et al.*¹ suggest that targeting SUMO1 could be an attractive approach to modifying specific pathogenic molecular pathways in heart failure. Although the results look promising, some caution is warranted. SUMO1 is

ubiquitous in its expression and biochemical function, and many of the heart's regulatory proteins and transcription factors are known to be SUMOylated^{9,13}, which could interfere with target specificity. Dysregulation of SUMO conjugation can contribute to the progression

of some human cancers⁶. Also, SUMO1 might bind to other proteins that modulate SERCA2a either directly or indirectly, leading to unwanted changes in Ca^{2+} handling^{9,13}. Nonetheless, therapeutic targeting of SUMO1 to treat heart failure remains an attractive proposition and should lead to new areas of investigation⁵. ■

Sudha K. Shenoy and Howard A. Rockman are in the Departments of Medicine, Cell Biology, and Molecular Genetics and Microbiology, Duke University Medical Center, Durham, North Carolina 27710, USA. e-mails: sudha@receptor-biol.duke.edu; h.rockman@duke.edu

1. Kho, C. *et al.* *Nature* **477**, 601–605 (2011).
2. Bers, D. M. *Nature* **415**, 198–205 (2002).
3. Cheng, H. & Lederer, W. J. *J. Physiol. Rev.* **88**, 1491–1545 (2008).
4. Rockman, H. A., Koch, W. J. & Lefkowitz, R. J. *Nature* **415**, 206–212 (2002).
5. Shah, A. M. & Mann, D. L. *Lancet* **378**, 704–712 (2011).
6. Geiss-Friedlander, R. & Melchior, F. *Nature Rev. Mol. Cell Biol.* **8**, 947–956 (2007).
7. Yeh, E. T. H. *J. Biol. Chem.* **284**, 8223–8227 (2009).
8. Matunis, M. J. & Pickart, C. M. *Nature Struct. Mol. Biol.* **12**, 565–566 (2005).
9. Wang, J. & Schwartz, R. J. *Circulation Res.* **107**, 19–29 (2010).
10. Plant, L. D. *et al.* *Proc. Natl Acad. Sci. USA* **107**, 10743–10748 (2010).
11. Benson, M. D. *et al.* *Proc. Natl Acad. Sci. USA* **104**, 1805–1810 (2007).
12. Stade, K. *et al.* *J. Biol. Chem.* **277**, 49554–49561 (2002).
13. Shenoy, S. K. & Lefkowitz, R. J. *Trends Pharmacol. Sci.* **32**, 521–533 (2011).

H.A.R. declares competing financial interests. See go.nature.com/dj5xzk for details.

CARBON CYCLE

A dent in carbon's gold standard

The global uptake of carbon by land plants may be greater than previously thought, according to observations based on the enigmatic Keeling curve of rising atmospheric carbon dioxide. SEE LETTER P.579

MATTHIAS CUNTZ

Estimates of how much carbon is taken up each year by the world's land plants are derived mainly from models of the carbon cycle. Worldwide measurements of terrestrial carbon exchange have yielded an estimate¹ of this global carbon uptake as 123 ± 8 petagrams carbon per year (Pg C yr^{-1} ; $1 \text{ Pg is } 10^{15} \text{ g}$). This is so close to earlier estimates derived from models and biomass production that 120 Pg C yr^{-1} can be taken

as carbon's 'gold standard'. But Welp and colleagues² remind us, on page 579 of this issue, that we should not be complacent — land ecosystems might be taking in considerably more carbon than we thought.

Our atmosphere is a perfect blender. Changes in its levels of trace gases — such as carbon dioxide — reveal variations in the total influx and uptake of its constituents. So if you measure the carbon exchange of a forest ecosystem, for example, you get the net exchange of all the carbon taken up by the trees

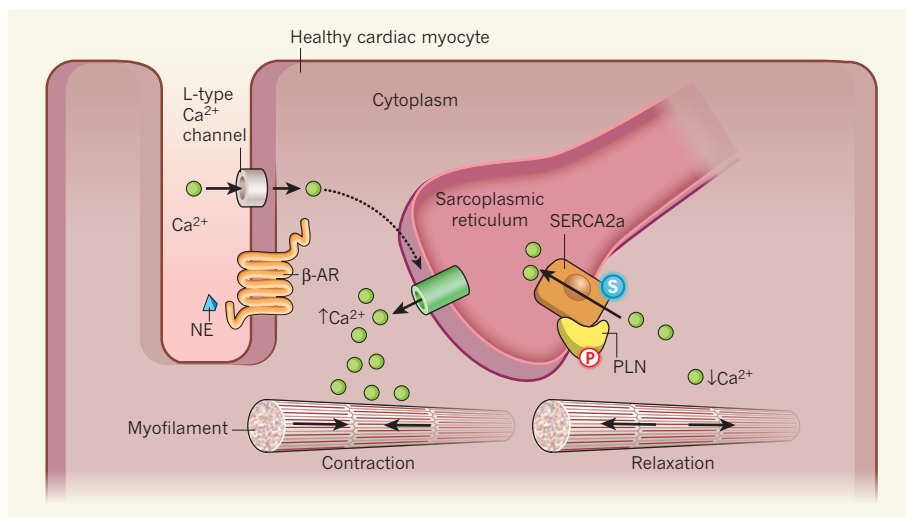


Figure 1 | The SERCA2a calcium pump and its modification by SUMO1. In a healthy cardiac myocyte, noradrenaline (NE) released from sympathetic nerves stimulates β -adrenergic receptors (β -AR) to enhance the activity of L-type calcium-ion (Ca^{2+}) channels on the cell membrane, leading to Ca^{2+} entry into the cell. This in turn triggers the release of a large amount of Ca^{2+} from the sarcoplasmic reticulum into the cytoplasm, and leads to myofilament contraction. Noradrenaline stimulation also causes phosphorylation (P) of phospholamban (PLN), releasing its inhibitory effect on the sarcoplasmic-reticulum Ca^{2+} -ATPase pump, SERCA2a. Activated SERCA2a binds Ca^{2+} ions from the cytoplasm and pumps them back into the sarcoplasmic reticulum, allowing the myofilaments to relax. Kho *et al.*¹ show that only SERCA2a modified by SUMO1 (S) can transport Ca^{2+} ions and maintain a healthy heart. In a failing heart (not shown), SUMO1 is depleted and SERCA2a remains in the un-SUMOylated form. It therefore cannot bind ATP and cannot effectively pump Ca^{2+} back into the sarcoplasmic-reticulum lumen, so relaxation of cardiac muscle is impaired.

that are known to be mutated in patients with familial dilated cardiomyopathy and those with conduction-system disease, suggesting that these disorders may also be attributable to defective SUMOylation⁹.

However, SUMO1 deficiency also causes degradation of the transcription factor SRF by caspase enzymes⁹, which are involved in programmed cell death. Therefore, restoring SUMO1 activity by introducing the SUMO1 transgene may block this and other cell-death pathways in the failing heart and lead to the observed recovery. Further investigation is needed to determine the cause of the reduced SUMO1 levels seen in heart failure: the protein does not undergo degradation and there is no evidence that gene transcription is inhibited.

In contrast to the stimulation of SERCA2a by SUMO1, some potassium-ion channels in the cell membrane are inhibited by SUMOylation^{10,11}; they regain their ion-selective activity when SUMOylation is blocked or reversed. Furthermore, there is evidence that SUMOylation could be critical for transport processes into the nucleus^{6,12}. It might therefore also facilitate Ca^{2+} transport across membranes, for example by imparting structural features of an 'on/off' switch to the ion channels.

Kho *et al.*¹ suggest that targeting SUMO1 could be an attractive approach to modifying specific pathogenic molecular pathways in heart failure. Although the results look promising, some caution is warranted. SUMO1 is

ubiquitous in its expression and biochemical function, and many of the heart's regulatory proteins and transcription factors are known to be SUMOylated^{9,13}, which could interfere with target specificity. Dysregulation of SUMO conjugation can contribute to the progression

of some human cancers⁶. Also, SUMO1 might bind to other proteins that modulate SERCA2a either directly or indirectly, leading to unwanted changes in Ca^{2+} handling^{9,13}. Nonetheless, therapeutic targeting of SUMO1 to treat heart failure remains an attractive proposition and should lead to new areas of investigation⁵. ■

Sudha K. Shenoy and Howard A. Rockman are in the Departments of Medicine, Cell Biology, and Molecular Genetics and Microbiology, Duke University Medical Center, Durham, North Carolina 27710, USA. e-mails: sudha@receptor-biol.duke.edu; h.rockman@duke.edu

1. Kho, C. *et al.* *Nature* **477**, 601–605 (2011).
2. Bers, D. M. *Nature* **415**, 198–205 (2002).
3. Cheng, H. & Lederer, W. J. *J. Physiol. Rev.* **88**, 1491–1545 (2008).
4. Rockman, H. A., Koch, W. J. & Lefkowitz, R. J. *Nature* **415**, 206–212 (2002).
5. Shah, A. M. & Mann, D. L. *Lancet* **378**, 704–712 (2011).
6. Geiss-Friedlander, R. & Melchior, F. *Nature Rev. Mol. Cell Biol.* **8**, 947–956 (2007).
7. Yeh, E. T. H. *J. Biol. Chem.* **284**, 8223–8227 (2009).
8. Matunis, M. J. & Pickart, C. M. *Nature Struct. Mol. Biol.* **12**, 565–566 (2005).
9. Wang, J. & Schwartz, R. J. *Circulation Res.* **107**, 19–29 (2010).
10. Plant, L. D. *et al.* *Proc. Natl Acad. Sci. USA* **107**, 10743–10748 (2010).
11. Benson, M. D. *et al.* *Proc. Natl Acad. Sci. USA* **104**, 1805–1810 (2007).
12. Stade, K. *et al.* *J. Biol. Chem.* **277**, 49554–49561 (2002).
13. Shenoy, S. K. & Lefkowitz, R. J. *Trends Pharmacol. Sci.* **32**, 521–533 (2011).

H.A.R. declares competing financial interests. See go.nature.com/dj5xzk for details.

CARBON CYCLE

A dent in carbon's gold standard

The global uptake of carbon by land plants may be greater than previously thought, according to observations based on the enigmatic Keeling curve of rising atmospheric carbon dioxide. SEE LETTER P.579

MATTHIAS CUNTZ

Estimates of how much carbon is taken up each year by the world's land plants are derived mainly from models of the carbon cycle. Worldwide measurements of terrestrial carbon exchange have yielded an estimate¹ of this global carbon uptake as 123 ± 8 petagrams carbon per year (Pg C yr^{-1} ; $1 \text{ Pg is } 10^{15} \text{ g}$). This is so close to earlier estimates derived from models and biomass production that 120 Pg C yr^{-1} can be taken

as carbon's 'gold standard'. But Welp and colleagues² remind us, on page 579 of this issue, that we should not be complacent — land ecosystems might be taking in considerably more carbon than we thought.

Our atmosphere is a perfect blender. Changes in its levels of trace gases — such as carbon dioxide — reveal variations in the total influx and uptake of its constituents. So if you measure the carbon exchange of a forest ecosystem, for example, you get the net exchange of all the carbon taken up by the trees

for photosynthesis and all the carbon released by the trees and soils through respiration. These gross-exchange fluxes — photosynthesis and respiration — are much larger than the net ecosystem exchange that is actually measured. On the global scale, the net flux is only a few per cent of the gross fluxes. Because small changes in photosynthesis and respiration can have big consequences for the net carbon uptake of terrestrial ecosystems, the interplay between photosynthesis and respiration must be well described in carbon-cycle models if they are to reliably project into the future. It is, however, almost impossible to measure individual components on scales larger than the size of a leaf, let alone on a regional or continental scale.

This is where Welp *et al.*² take advantage of the composition of oxygen isotopes in CO₂ — the chemical signature of which changes if one ¹⁶O oxygen atom in CO₂ is replaced by a heavier ¹⁸O atom. Carbon dioxide dissolves in water and exchanges its oxygen with water's oxygen to equilibrium, so CO₂ is tagged by the water it comes into contact with. Different waters have distinct isotopic compositions owing to evaporation processes in soils and leaves — the lighter molecules evaporate faster, and the heavier ones fall behind. As a result, the oxygen isotopic composition in CO₂ is very sensitive to photosynthesis and respiration: more photosynthesis means more ¹⁸O, and hence higher oxygen-isotope ratios in the atmosphere.

Using an impressive 30-year record of the isotopic composition of atmospheric CO₂, Welp *et al.*² assess the mean atmospheric residence time for oxygen atoms in CO₂. Their 11 time series were started in the 1970s by the late Charles Keeling, alongside the famous record of total atmospheric CO₂ at Mauna Loa in Hawaii (Fig. 1). Welp *et al.* identified a strong correlation between the observed inter-annual variability of the oxygen isotopes and the El Niño–Southern Oscillation (ENSO). Such a correlation has previously been established for the isotopic composition of water³ and, consequently, is now found in the oxygen isotopes of CO₂ as well. From the mean residence time of the oxygen atoms in CO₂,

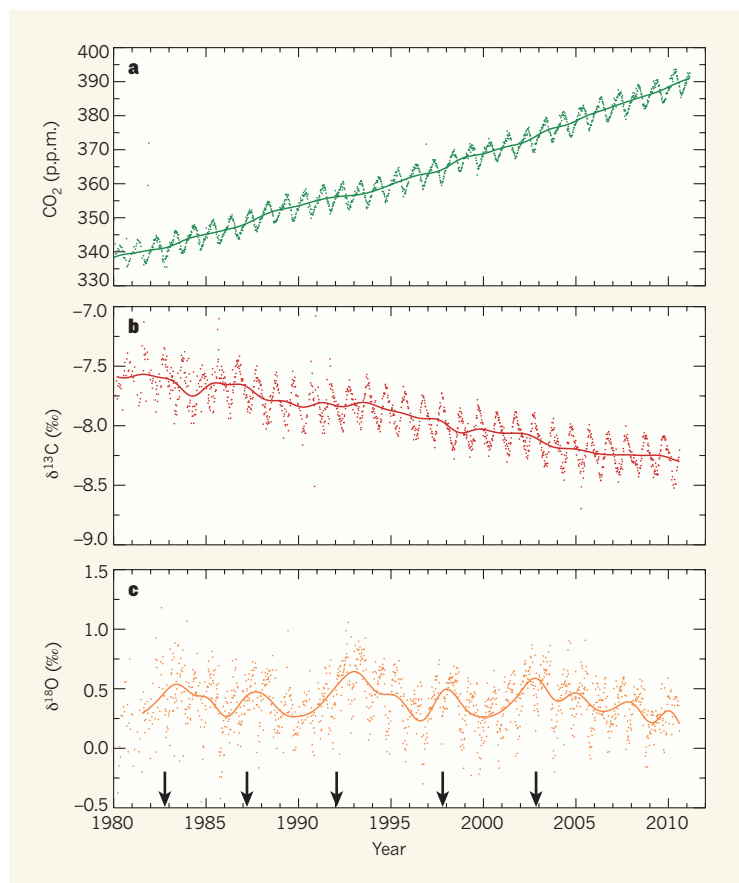


Figure 1 | Atmospheric CO₂ concentrations and isotope composition measured at Mauna Loa, Hawaii. **a**, Keeling curve of CO₂ concentrations since the 1980s (ref. 7). Dots are single measurements or daily averages; line indicates the long-term trend. **b**, Carbon-isotope composition of the CO₂ at Mauna Loa. Here, $\delta^{13}\text{C}$ is the deviation of the ¹³C/¹²C ratio from a standard value. Because the carbon cycle is the major influence on both CO₂ concentrations and ¹³C/¹²C ratios, the curves in **a** and **b** correlate well with each other (that is, the downward trend in **b** mirrors the upward trend in **a** and so do the seasonal variations). **c**, The oxygen-isotope composition of the CO₂ is influenced not only by the carbon cycle, but also by the water cycle, and so does not correlate simply with CO₂ concentration; $\delta^{18}\text{O}$ is the deviation of the ¹⁸O/¹⁶O ratio from a standard value. Welp and colleagues² find that the interannual variations in $\delta^{18}\text{O}$ correlate with the El Niño–Southern Oscillation (arrows indicate El Niño events). Their analysis of the oxygen-isotope data also provides a new estimate of global carbon uptake on land. p.p.m., parts per million. (Data are publicly available on the Scripps Institution of Oceanography website⁸.)

Welp *et al.* arrive at a best guess of global productivity of 150–175 Pg C yr⁻¹ — some 25–45% more than the gold standard.

This inference hinges on a set of assumptions and estimates. It depends, for example, on how many CO₂ molecules actually enter a plant before one molecule is fixed by photosynthesis. The authors think that plants eventually fix some 43% of all CO₂ molecules entering a leaf; however, if this were only 34%, the isotope-based estimate would fall to about 120 Pg C yr⁻¹, the current gold standard.

The global value of 43% is derived from carbon-cycle models and remains uncertain, because it depends on the details of the models' formulation. It also depends on the distribution of different plant types. For example, some savannah grasses and maize (corn)

fix carbon more efficiently through the C₄ metabolic pathway, rather than by the usual C₃ route, thereby enabling them to fix about 60% of the CO₂ molecules that enter the plant. Hence, the global abundance and distribution of C₄ plants are important in estimates of global productivity, whether these are derived from modelling, actual measurements or isotope-composition data. One carbon-cycle model, for example, increased global productivity by more than 20% simply by substituting a new map of C₄-plant distribution⁴.

So it looks as though we are stuck with model-based estimates that are hard to validate globally. But other isotopes might yet come to the rescue: the isotopic composition of the carbon atoms in CO₂ provides a measure of the percentage of carbon that is fixed⁵. This could constrain estimates such as that offered by Welp *et al.*, but it could also constrain C₄-plant distribution and therefore help non-isotopic estimates of global production as well. And the carbon-isotope estimate of the percentage of carbon that is fixed might be further refined with the help of carbonyl sulphide, a new tracer of leaves' ability to take up CO₂ (ref. 6).

Gold does not tarnish easily. With their approach, and by making their long-term records publicly available, Welp and colleagues² are preparing the ground to combine

these pieces of information and polish up carbon's gold standard of the future. ■

Matthias Cuntz is at the UFZ – Helmholtz Centre for Environmental Research, Permoserstrasse 15, 04318 Leipzig, Germany. e-mail: matthias.cuntz@ufz.de

1. Beer, C. *et al.* *Science* **329**, 834–838 (2010).
2. Welp, L. R. *et al.* *Nature* **477**, 579–582 (2011).
3. Werner, M. & Heimann, M. *J. Geophys. Res.* <http://dx.doi.org/10.1029/2001JD900253> (2002).
4. Still, C. J. *et al.* *Glob. Biogeochem. Cycles* <http://dx.doi.org/10.1029/2001GB001807> (2003).
5. Farquhar, G. D., O'Leary, M. H. & Berry, J. A. *Aust. J. Plant Physiol.* **9**, 121–137 (1982).
6. Stimler, K. *et al.* *New Phytol.* **186**, 869–878 (2010).
7. Keeling, C. D. *et al.* in *A History of Atmospheric CO₂ and Its Effects on Plants, Animals, and Ecosystems* (eds Ehleringer, J. R., Cerling, T. E. & Dearing, M. D.) 83–113 (Springer, 2005).
8. http://scrippsco2.ucsd.edu/data/atmospheric_co2.html

Crystal structure of the β_2 adrenergic receptor–Gs protein complex

Søren G. F. Rasmussen^{1,2*}, Brian T. DeVree^{3*}, Yaozhong Zou¹, Andrew C. Kruse¹, Ka Young Chung¹, Tong Sun Kobilka¹, Foon Sun Thian¹, Pil Seok Chae⁴, Els Pardon^{5,6}, Diane Calinski³, Jesper M. Mathiesen¹, Syed T. A. Shah⁷, Joseph A. Lyons⁷, Martin Caffrey⁷, Samuel H. Gellman⁴, Jan Steyaert^{5,6}, Georgios Skiniotis⁸, William I. Weiss^{1,9}, Roger K. Sunahara³ & Brian K. Kobilka¹

G protein-coupled receptors (GPCRs) are responsible for the majority of cellular responses to hormones and neurotransmitters as well as the senses of sight, olfaction and taste. The paradigm of GPCR signalling is the activation of a heterotrimeric GTP binding protein (G protein) by an agonist-occupied receptor. The β_2 adrenergic receptor (β_2 AR) activation of Gs, the stimulatory G protein for adenylyl cyclase, has long been a model system for GPCR signalling. Here we present the crystal structure of the active state ternary complex composed of agonist-occupied monomeric β_2 AR and nucleotide-free Gs heterotrimer. The principal interactions between the β_2 AR and Gs involve the amino- and carboxy-terminal α -helices of Gs, with conformational changes propagating to the nucleotide-binding pocket. The largest conformational changes in the β_2 AR include a 14 Å outward movement at the cytoplasmic end of transmembrane segment 6 (TM6) and an α -helical extension of the cytoplasmic end of TM5. The most surprising observation is a major displacement of the α -helical domain of Gs relative to the Ras-like GTPase domain. This crystal structure represents the first high-resolution view of transmembrane signalling by a GPCR.

Introduction

The β_2 adrenergic receptor (β_2 AR) has been a model system for the large and diverse family of G protein-coupled receptors (GPCRs) for over 40 years. It was one of the first GPCRs to be characterized by radioligand binding, and it was the first neurotransmitter receptor to be cloned¹ and structurally determined by crystallography^{2,3}. The β_2 AR was initially identified based on its physiological and pharmacological properties, but it was not known if receptors and G proteins were separate entities, or parts of the same protein⁴. Subsequent biochemical studies led to the isolation and purification of functional β_2 AR and Gs, the stimulatory G protein that activates adenylyl cyclase, and the reconstitution of this signalling complex in phospholipid vesicles^{5,6}. The cooperative interactions of β_2 AR and Gs observed in ligand binding assays formed the foundation of the ternary complex model of GPCR activation^{7,8}. In the ternary complex consisting of agonist, receptor and G protein, the affinity of the receptor for agonist is enhanced and the specificity of the G protein for guanine nucleotides changes in favour of GTP over GDP. The GPCR field has evolved markedly since these initial studies. Isolation of the genes and cDNAs for the β_2 AR and other GPCRs using protein sequencing and expression cloning led to the expansion of the family by homology cloning. More recently, sequencing of the human genome led to the identification of over 800 GPCR genes⁹. Experimental tools for identifying protein–protein interactions and for expression and silencing of genes have revealed a complex network of cellular signalling and regulatory pathways including G protein-independent activation of cytosolic kinases^{10,11}. Nevertheless, the β_2 AR continues to be a relevant model for most aspects of GPCR pharmacology, signalling and regulation.

Notwithstanding the remarkable advances in this field, we still know relatively little about the structural basis for transmembrane signalling by GPCRs. Figure 1 shows the G protein cycle for the β_2 AR–Gs complex. Agonist binding to the β_2 AR promotes interactions with GDP-bound Gs $\alpha\beta\gamma$ heterotrimer, leading to the exchange of GDP for GTP, and the functional dissociation of Gs into G α -GTP and G $\beta\gamma$ subunits. The separate G α -GTP and G $\beta\gamma$ subunits can modulate the activity of different cellular effectors (channels, kinases or other enzymes). The intrinsic GTPase activity of G α s leads to hydrolysis of GTP to GDP and the reassociation of G α -GDP and G $\beta\gamma$ subunits, and the termination of signalling. The active state of a GPCR can be defined as that conformation that couples to and stabilizes a nucleotide-free G protein. In this agonist– β_2 AR–Gs ternary complex, Gs has a higher affinity for GTP than GDP, and the β_2 AR has an approximately 100-fold higher affinity for agonists than does β_2 AR alone. In an effort to understand the structural basis for GPCR signalling, we crystallized the β_2 AR–Gs complex.

Crystallization of the β_2 AR–Gs complex

The first challenge for crystallogenes was to prepare a stable β_2 AR–Gs complex in detergent solution. The β_2 AR and Gs couple efficiently in lipid bilayers, but not in detergents used to solubilize and purify these proteins. We found that a relatively stable β_2 AR–Gs complex could be prepared by mixing purified GDP-Gs (approximately 100 μ M final concentration) with a molar excess of purified β_2 AR bound to a high affinity agonist (BI-167107, Boehringer Ingelheim)¹² in dodecylmaltoside solution. Apyrase, a non-selective purine pyrophosphatase, was added to hydrolyse GDP released from Gs on forming a complex with the β_2 AR. Removal of GDP was essential because both GDP and GTP

¹Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford, California 94305, USA. ²Department of Neuroscience and Pharmacology, The Panum Institute, University of Copenhagen, 2200 Copenhagen N, Denmark. ³Department of Pharmacology, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA. ⁴Department of Chemistry, University of Wisconsin, Madison, Wisconsin 53706, USA. ⁵Department of Molecular and Cellular Interactions, Vlaams Instituut voor Biotechnologie (VIB), Vrije Universiteit Brussel, B-1050 Brussel, Belgium. ⁶Structural Biology Brussels, Vrije Universiteit Brussel, B-1050 Brussels, Belgium. ⁷Membrane Structural and Functional Biology Group, Schools of Medicine and Biochemistry & Immunology, Trinity College, Dublin 2, Ireland. ⁸Life Sciences Institute and Department of Biological Chemistry, University of Michigan, Ann Arbor, Michigan 48109, USA. ⁹Department of Structural Biology, Stanford University School of Medicine, Stanford, California 94305, USA.

*These authors contributed equally to this work.

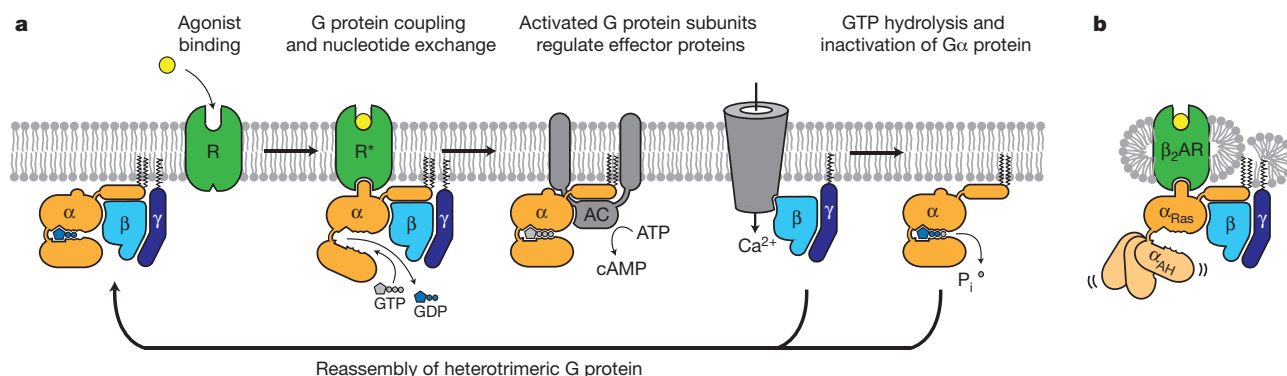


Figure 1 | G protein cycle for the β_2 AR-Gs complex. **a**, Extracellular agonist binding to the β_2 AR leads to conformational rearrangements of the cytoplasmic ends of transmembrane segments that enable the Gs heterotrimer (α , β , and γ) to bind the receptor. GDP is released from the α subunit upon formation of β_2 AR-Gs complex. The GTP binds to the nucleotide-free α subunit resulting in dissociation of the α and $\beta\gamma$ subunits from the receptor. The subunits regulate their respective effector proteins adenylyl cyclase (AC) and Ca^{2+} channels. The Gs heterotrimer reassembles from α and $\beta\gamma$ subunits following hydrolysis of GTP to GDP in the α subunit. **b**, The purified nucleotide-free β_2 AR-Gs protein complex maintained in detergent micelles. The $G\alpha_s$ subunit consists of two domains, the Ras domain (α_{Ras}) and the α -helical domain (α_{AH}). Both are involved in nucleotide binding. In the nucleotide-free state, the α_{AH} domain has a variable position relative the α_{Ras} domain.

can disrupt the high-affinity interaction between β_2 AR and Gs (Supplementary Fig. 1a). The complex was subsequently purified by sequential antibody affinity chromatography and size-exclusion chromatography. The stability of the complex was enhanced by exchanging it into a recently developed maltose neopentyl glycol detergent MNG-3 (NG310, Anatrace)¹³. The complex could be incubated at room temperature for 24 h without any noticeable degradation; however, initial efforts to crystallize the complex using sparse matrix screens in detergent micelles, bicelles and lipidic cubic phase (LCP) failed.

To further assess the quality of the complex, we analysed the protein by single particle electron microscopy³⁴. The results confirmed that the complex was monodisperse, but revealed two potential problems for obtaining diffraction of quality crystals. First, the detergent used to stabilize the complex formed a large micelle, leaving little polar surface on the extracellular side of the β_2 AR-Gs complex for the formation of crystal lattice contacts. Our initial approach to this problem, which was to generate antibodies to the extracellular surface, was not successful. As an alternative approach, we replaced the unstructured amino terminus of the β_2 AR with T4 lysozyme (T4L). We previously used T4L to facilitate crystallography of the inactive β_2 AR by inserting T4L between the cytoplasmic ends of TM5 and TM6 (ref. 3). Several different amino-terminal fusion proteins were prepared and single particle electron microscopy was used to identify a fusion with a relatively fixed orientation of T4L in relation to the β_2 AR.

The second problem revealed by single particle electron microscopy analysis was increased variability in the positioning of the α -helical component of the $G\alpha_s$ subunit. $G\alpha_s$ consists of two domains, the Ras-like GTPase domain ($G\alpha_{\text{Ras}}$), which interacts with the β_2 AR and the $G\beta$ subunit, and the α -helical domain ($G\alpha_{\text{AH}}$)¹⁴. The interface of the two $G\alpha_s$ subdomains forms the nucleotide-binding pocket (Fig. 1), and electron microscopy two-dimensional (2D) averages and three-dimensional (3D) reconstructions show that in the absence of guanine nucleotide, $G\alpha_{\text{AH}}$ has a variable position relative to the complex of T4L- β_2 AR- $G\alpha_{\text{Ras}}$ - $G\beta\gamma$ (Fig. 1b)³⁴.

We attributed the variable position of $G\alpha_{\text{AH}}$ to the empty nucleotide-binding pocket. However, as noted above both GDP and non-hydrolysable GTP analogues disrupt the β_2 AR-Gs complex (Supplementary Fig. 1). The addition of the pyrophosphate analogue phosphonoformate (fosfarnet) led to a significant increase in stabilization of $G\alpha_{\text{AH}}$ as determined by electron microscopy analysis of the detergent-solubilized complex³⁴. Crystallization trials were carried out in LCP using a modified monolein (7.7 MAG, see Methods) designed to accommodate the large hydrophilic component of the T4L- β_2 AR-Gs complex¹⁵. Although we were able to obtain small crystals that

and Ca^{2+} channels. The Gs heterotrimer reassembles from α and $\beta\gamma$ subunits following hydrolysis of GTP to GDP in the α subunit. **b**, The purified nucleotide-free β_2 AR-Gs protein complex maintained in detergent micelles. The $G\alpha_s$ subunit consists of two domains, the Ras domain (α_{Ras}) and the α -helical domain (α_{AH}). Both are involved in nucleotide binding. In the nucleotide-free state, the α_{AH} domain has a variable position relative the α_{Ras} domain.

diffracted to 7 Å, we were unable to improve their quality through the use of additives and other modifications.

In an effort to generate an antibody that would further stabilize the complex and facilitate crystallography, we crosslinked β_2 AR and the Gs heterotrimer with a small, homobifunctional amine-reactive cross-linker and used this stabilized complex to immunize llamas. Llamas and other camelids produce antibodies devoid of light chains. The single domain antigen binding fragments of these heavy-chain-only antibodies, known as nanobodies, are small (15 kDa), rigid, and are easily cloned and expressed in *Escherichia coli* (Methods)¹⁶. We obtained a nanobody (Nb35) that binds to the complex and prevents dissociation of the complex by GTP γ S (Supplementary Fig. 1). The T4L- β_2 AR-Gs-Nb35 complex was used to obtain crystals that grew to 250 μm (Supplementary Fig. 2) in LCP (7.7 MAG) and diffracted to 2.9 Å. A 3.2 Å data set was obtained from 20 crystals and the structure was determined by molecular replacement (Methods).

The β_2 AR-Gs complex crystallized in primitive monoclinic space group $P2_1$, with a single complex in each asymmetric unit. Figure 2a shows the crystallographic packing interactions. Complexes are arrayed in alternating aqueous and lipidic layers with lattice contacts formed almost exclusively between soluble components of the complex, leaving receptor molecules suspended between G protein layers and widely separated from one another in the plane of the membrane. Extensive lattice contacts are formed among all the soluble proteins, probably accounting for the strong overall diffraction and remarkably clear electron density for the G protein. Nb35 and T4L facilitated crystal formation. Nb35 packs at the interface of the $G\beta$ and $G\alpha$ subunits, with the complementarity determining region (CDR) 1 interacting primarily with $G\beta$ and a long CDR3 loop interacting with both $G\beta$ and $G\alpha$ subunits. The framework regions of Nb35 from one complex also interact with $G\alpha$ subunits from two adjacent complexes. T4L is linked to the β_2 AR only through amino-terminal fusion, but packs against the amino terminus of the $G\beta$ subunit of one complex, the carboxy terminus of the $G\gamma$ subunit of another complex, and the $G\alpha$ subunit of yet another complex. Figure 2b shows the structure of the complete complex including T4L and Nb35, and Fig. 2c shows the β_2 AR-Gs complex alone.

Structure of the active-state β_2 AR

The β_2 AR-Gs structure provides the first high-resolution insight into the mechanism of signal transduction across the plasma membrane by a GPCR, and the structural basis for the functional properties of the ternary complex. Figure 3a compares the structures of the agonist-bound receptor in the β_2 AR-Gs complex and the inactive

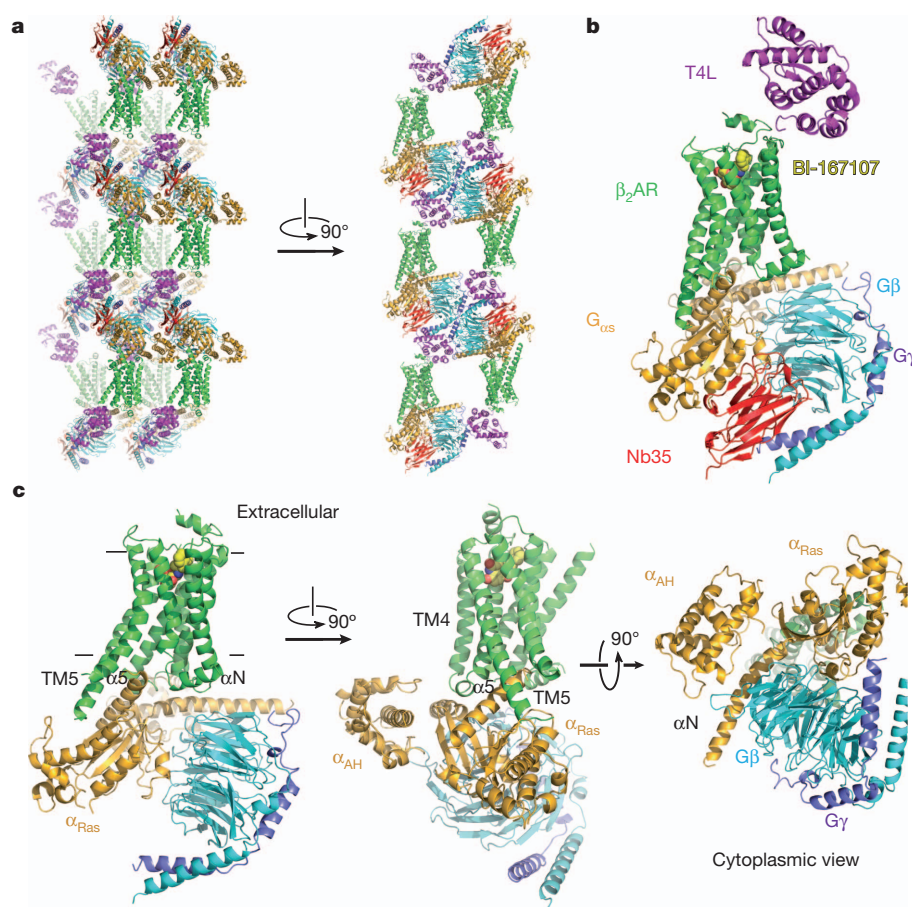


Figure 2 | Overall structure of the β_2 AR-Gs complex. **a**, Lattice packing of the complex shows alternating layers of receptor and G protein within the crystal. Abundant contacts are formed among proteins within the aqueous layers. **b**, The overall structure of the asymmetric unit contents shows the β_2 AR (green) bound to an agonist (yellow spheres) and engaged in extensive interactions with G α s (orange). G α s together with G β (cyan) and G γ (purple) constitute the heterotrimeric G protein Gs. A Gs-binding nanobody (red) binds the G protein between the α and β subunits. The nanobody (Nb35) facilitates crystallization, as does T4 lysozyme (magenta) fused to the amino terminus of the β_2 AR. **c**, The biological complex omitting crystallization aids, showing its location and orientation within a cell membrane.

carazolol-bound β_2 AR. The largest difference between the inactive and active structures is a 14 Å outward movement of TM6 when measured at the C α carbon of E268. There is a smaller outward movement and extension of the cytoplasmic end of the TM5 helix by 7 residues. A stretch of 26 amino acids in the third intracellular loop (ICL3) is disordered. Another notable difference between inactive and active structures is the second intracellular loop (ICL2), which forms an extended loop in the inactive β_2 AR structure and an α -helix in the β_2 AR-Gs complex. This helix is also observed in the β_2 AR-Nb80 structure (Fig. 3b); however, it may not be a feature that is unique to the active state, because it is also observed in the inactive structure of the highly homologous avian β_1 AR (ref. 17).

The quality of the electron density maps for the β_2 AR is highest at the β_2 AR-G α sRas interface, and much weaker for the extracellular half. The extracellular half of the receptor is not stabilized by any packing interactions either laterally with adjacent receptors in the membrane or through the extracellular surface. Instead, the extracellular region is indirectly tethered to the well-packed soluble components by the amino-terminal fusion to T4 lysozyme (Fig. 2a). Given the flexible and dynamic nature of GPCRs, the absence of stabilizing packing interactions may lead to structural heterogeneity in the extracellular half of the receptor and, consequently, to the limited quality of the electron density maps. However, the overall structure of the β_2 AR in the T4L- β_2 AR-Gs complex is very similar to our recent active-state structure of β_2 AR stabilized by a G protein mimetic nanobody (Nb80)¹². In the β_2 AR-Nb80 crystal, each receptor molecule has extensive packing interactions with adjacent receptors and the quality of the electron density maps for the agonist-bound β_2 AR in this complex is remarkably good for a 3.5 Å structure. Therefore, the β_2 AR-Nb80 structure allows us to confidently model BI-167107 here, and provide a more reliable view of the conformational rearrangements of amino acids around the ligand-binding pocket and between the ligand-binding pocket and the Gs-coupling interface¹².

The overall root mean square deviation between the β_2 AR components in the β_2 AR-Gs and β_2 AR-Nb80 structures is approximately 0.6 Å, and they differ most at the cytoplasmic ends of transmembrane helices 5 and 6 where they interact with the different proteins (Fig. 3b–d). The largest divergence is a 3 Å outward movement at the end of helix 6 in the β_2 AR-Gs complex. However, the differences between these two structures are very small at the level of the most highly conserved amino acids (E/DRY and NPxxY), which are located at the cytoplasmic ends of the transmembrane segments (Fig. 3c, d). These conserved sequences have been proposed to be important for activation or for maintaining the receptor in the inactive state¹⁸. Of these residues, only Arg 131 differs significantly between these two structures. In β_2 AR-Nb80 Arg 131 interacts with Nb80, whereas in the β_2 AR-Gs structure Arg 131 packs against Tyr 391 of G α s (Supplementary Fig. 3). The high structural similarity is in agreement with the functional similarity of these two proteins. The β_2 AR-Nb80 complex shows the same high affinity for the agonist isoproterenol as does the β_2 AR-Gs complex¹², consistent with high structural homology around the ligand binding pocket.

The active state of the β_2 AR is stabilized by extensive interactions with G α sRas (Fig. 4). There are no direct interactions with G β or G γ subunits. The total buried surface of the β_2 AR-G α sRas interface is 2,576 Å² (1,300 Å² for G α sRas and 1,276 Å² for the β_2 AR). This interface is formed by ICL2, TM5 and TM6 of the β_2 AR, and by α 5-helix, the α N- β 1 junction, the top of the β 3-strand, and the α 4-helix of G α sRas (see Supplementary Table 1 for specific interactions). Some of the β_2 AR sequences involved in this interaction have been shown to have a role in G protein coupling; however, there is no clear consensus sequence for Gs-coupling specificity when these segments are aligned with other GPCRs. Perhaps this is not surprising considering that the β_2 AR also couples to Gi and that many GPCRs couple to more than one G protein isoform. Of the 21 amino acids of Gs that are within 4 Å of the β_2 AR, only five are identical between Gs and Gi, and all of these

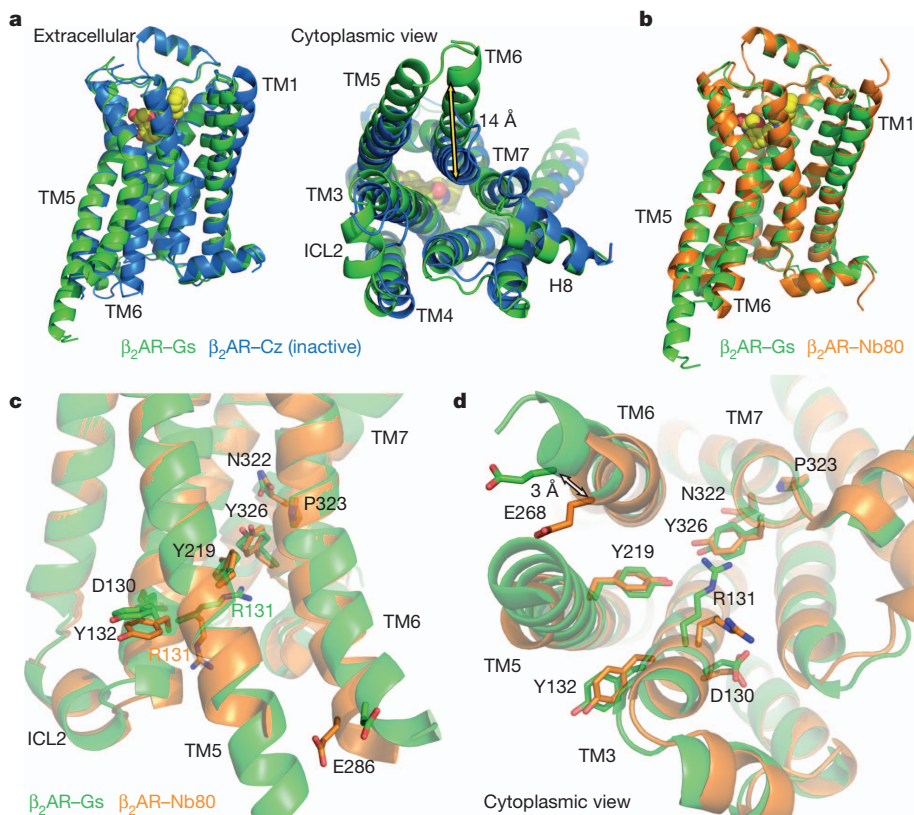


Figure 3 | Comparison of active and inactive β_2 AR structures. **a**, Side and cytoplasmic views of the β_2 AR-Gs structure (green) compared to the inactive carazolol-bound β_2 AR structure³ (blue). Significant structural changes are seen for the intracellular domains of TM5 and TM6. TM5 is extended by two helical turns whereas TM6 is moved outward by 14 Å as measured at the α -carbons of Glu 268 (yellow arrow) in the two structures. **b**, β_2 AR-Gs compared with the nanobody-stabilized active state β_2 AR-Nb80 structure¹² (orange). **c**, The positions of residues in the E/DRY and NPxxY motifs and other key residues of the β_2 AR-Gs and β_2 AR-Nb80 structures. All residues occupy very similar positions except Arg 131 which in the β_2 AR-Nb80 structure interacts with the nanobody. **d**, View from the cytoplasmic side of residues shown in **c**.

are in the carboxy-terminal α helix. The structural basis for G protein coupling specificity must therefore involve more subtle features of the secondary and tertiary structure. Nevertheless, a noteworthy interaction involves Phe 139, which is located at the beginning of the ICL2 helix and sits in a hydrophobic pocket formed by G α s His 41 at the beginning of the β 1-strand, Val 217 at the start of the β 3-strand and Phe 376, Cys 379, Arg 380 and Ile 383 in the α 5-helix (Fig. 4c). The β_2 AR mutant F139A has severely impaired coupling to Gs¹⁹. The residue corresponding to Phe 139 is a Phe or Leu on almost all Gs coupled receptors, but is more variable in GPCRs known to couple to other G proteins. Of interest, the ICL2 helix is stabilized by an interaction between Asp 130 of the conserved DRY sequence and Tyr 141

in the middle of the ICL2 helix (Fig. 4c). Tyr 141 has been shown to be a substrate for the insulin receptor tyrosine kinase²⁰; however, the functional significance of this phosphorylation is currently unknown.

The lack of direct interactions between the β_2 AR and G $\beta\gamma$ is somewhat unexpected given that a heterotrimer is required for efficient coupling to a GPCR. Whereas G β does not interact directly with the β_2 AR, it has an indirect but important role in coupling by stabilizing the amino-terminal α helix of G α s (Fig. 2c). Several models involving GPCR dimers propose that one of the protomers interacts predominantly with G α while the other interacts with G $\beta\gamma$ ^{21–23}. Consistent with these models, biochemical and biophysical evidence suggests that G α i2 forms a stable complex with a LTB4 receptor dimer²⁴.

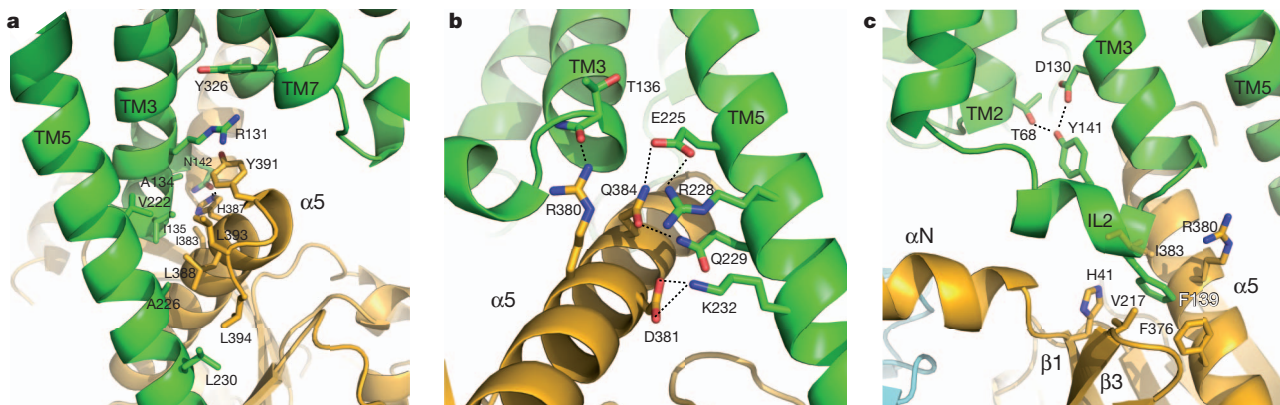


Figure 4 | Receptor-G protein interactions. **a**, **b**, The α 5-helix of G α s docks into a cavity formed on the intracellular side of the receptor by the opening of transmembrane helices 5 and 6. **a**, Within the transmembrane core, the interactions are primarily non-polar. An exception involves packing of Tyr 391 of the α 5-helix against Arg 131 of the conserved DRY sequence in TM3 (see also Supplementary Fig. 3). Arg 131 also packs against Tyr 326 of the conserved

NPxxY sequence in TM7. **b**, As α 5-helix exits the receptor it forms a network of polar interactions with TM5 and TM3. **c**, Receptor residues Thr 68 and Asp 130 interact with the ICL2 helix of the β_2 AR via Tyr 141, positioning the helix so that Phe 139 of the receptor docks into a hydrophobic pocket on the G protein surface, thereby structurally linking receptor-G protein interactions with the highly conserved DRY motif of the β_2 AR.

Whereas the β_2 AR efficiently activates Gs as a monomer, extensive biochemical and biophysical evidence supports the existence of β_2 AR dimers or oligomers in living cells²⁵. Therefore, we cannot exclude the possibility that in cell membranes one protomer of a β_2 AR dimer may interact with the G $\beta\gamma$ subunit.

Structure of activated Gs

The most surprising observation in the β_2 AR–Gs complex is the large displacement of the G α sAH relative to G α sRas (an approximately 127° rotation about the junction between the domains) (Fig. 5a). G α sAH moves as a rigid body as shown by the alignment of β_2 AR–Gs with G α sAH from the crystal structure of G α s–GTP γ S²⁶ (Supplementary Fig. 4). In the structure of G α s–GTP γ S, the nucleotide-binding pocket is formed by the interface between G α sRas and G α sAH. Guanine nucleotide binding stabilizes the interaction between these two domains. The loss of this stabilizing effect of guanine nucleotide binding is consistent with the high flexibility observed for G α sAH in single particle electron microscopy analysis of the detergent-solubilized complex³⁴. It is also in agreement with the increase in deuterium exchange at the interface between these two domains upon formation of the complex³⁵. Recently double electron-electron resonance (DEER) spectroscopy was used to document large (up to 20 Å) changes in distance between nitroxide probes positioned on the Ras and α -helical domains of Gi upon formation of a complex with light-activated rhodopsin²⁷. Finally, it has been shown that G α sRas and G α sAH can form a functional GTPase when expressed together as separate proteins²⁸. Therefore, it is perhaps not surprising that G α sAH is displaced relative to G α sRas; however, its location in this crystal structure most probably reflects only one of an ensemble of conformations that it can adopt under physiological conditions, but has been stabilized by crystal packing interactions (Supplementary Fig. 5).

A potential concern is that Nb35, which was used to facilitate crystallization, may be responsible for the displacement of the G α sAH. However, Nb35, which binds at the interface between G β and G α sRas (Fig. 2b and Supplementary Fig. 6), would not be expected to interact with the G α sAH or interfere with its interactions with G α sRas. None of the Nb35 contacts on the Ras domain are involved in interactions with G α sAH on the basis of the crystal structure of G α s–GTP γ S (1AZT). Moreover, if we superimpose the structures of the Ras domains of G α s–GTP γ S (1AZT) and β_2 AR–Gs, there is no overlap between Nb35 and the α -helical domain of G α s–GTP γ S (Supplementary Fig. 6). Similarly, if we align the G β subunits of the Gi–GDP heterotrimer (1GP2) with

that of β_2 AR–Gs, there is no overlap between Nb35 and the α -helical domain of Gi (Supplementary Fig. 6). This analysis is in agreement with single particle electron microscopy studies which provide further evidence that Nb35 does not disrupt interactions between G α sAH and G α sRas³⁴.

The conformational links between the β_2 AR and the nucleotide-binding pocket primarily involve the amino- and carboxy-terminal helices of G α s (Fig. 4). Figure 5b focuses on the region of G α sRas that undergoes the largest conformational change when comparing the structure of G α sRas from the β_2 AR–Gs complex with that from the G α s–GTP γ S complex²⁶. The largest difference is observed for the α 5-helix, which is displaced 6 Å towards the receptor and rotated as the carboxy-terminal end projects into transmembrane core of the β_2 AR. Previous studies using a variety of approaches have demonstrated the important role of the α 5-helix in GPCR–G protein interactions^{29,30}. Associated with movement of the α 5-helix, the β 6– α 5 loop, which interacts with the guanine ring in the G α s–GTP γ S structure, is displaced outward, away from the nucleotide-binding pocket (Fig. 5b–d). The movement of the α 5-helix is also associated with changes in interactions between this helix and the β 6-strand, the α N– β 1 loop, and the α 1-helix. The β 1-strand forms another link between the β_2 AR and the nucleotide-binding pocket. The carboxy-terminal end of this strand changes conformation around Gly 47, and there are further changes in the β 1– α 1 loop (P-loop) that coordinates the β -phosphate in the GDP and GTP-bound forms (Fig. 5b–d). The observations in the crystal structure are in agreement with deuterium exchange experiments where there is enhanced deuterium exchange in the β 1-strand and the amino-terminal end of the α 5-helix upon formation of the nucleotide-free β_2 AR–Gs complex³⁵. The deuterium exchange-mass spectrometry (DXMS) studies provide additional insights into the dynamic nature of these conformational changes in Gs upon complex formation³⁵.

The structure of a GDP-bound Gs heterotrimer has not been determined, so it is not possible to directly compare the G α s–G $\beta\gamma$ interface before and after formation of the β_2 AR–Gs complex. On the basis of the structure of the GDP-bound Gi heterotrimer³¹, we do not observe large changes in interactions between G α sRas and G $\beta\gamma$ upon formation of the complex with β_2 AR. This is also consistent with deuterium exchange studies³⁵. As discussed above, Nb35 binds at the interface between G α sRas and G β (Fig. 2b); therefore, we cannot exclude the possibility that Nb35 may influence the relative orientation of the G α sRas–G $\beta\gamma$ interface in the crystal structure.

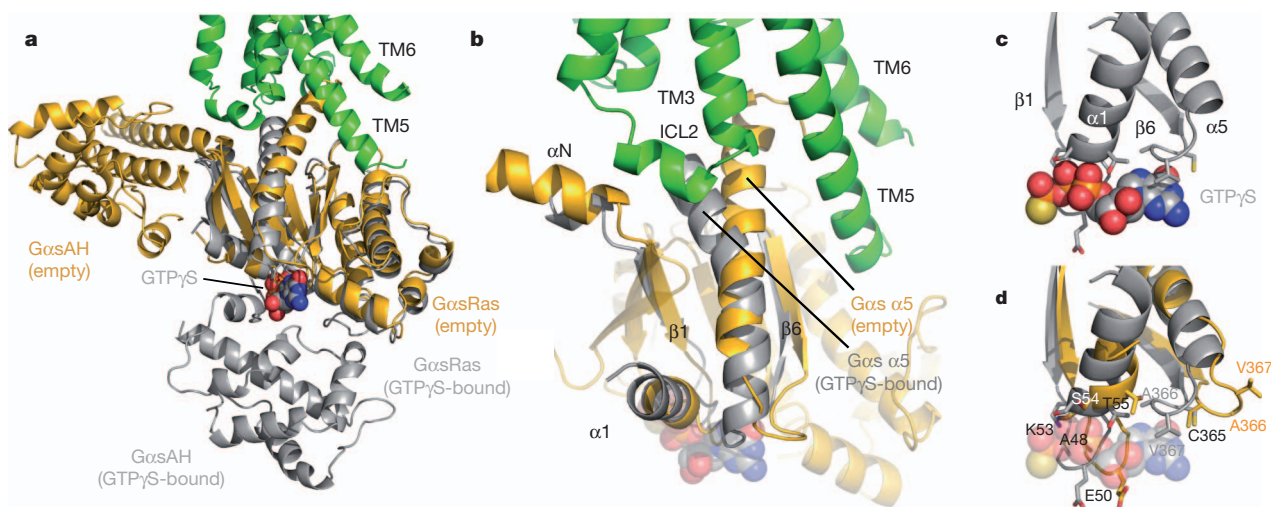


Figure 5 | Conformational changes in G α s. **a**, A comparison of G α s in the β_2 AR–Gs complex (orange) with the GTP γ S-bound G α s (grey)²⁶ (PDB ID: 1AZT). GTP γ S is shown as spheres. The helical domain of G α s (G α sAH) shows a marked displacement relative to its position in the GTP γ S-bound state. **b**, The α 5-helix of G α s is rotated and displaced towards the β_2 AR, perturbing the

β 6– α 5 loop, which otherwise forms part of the GTP γ S-binding pocket. **c**, The β 1– α 1 loop (P-loop) and β 6– α 5 loop of G α s interact with the phosphates and purine ring, respectively, of GTP γ S in the G α s–GTP γ S structure. **d**, The β 1– α 1 and β 6– α 5 loops are rearranged in the nucleotide-free β_2 AR–Gs structure.

Assembly of the β_2 AR–Gs complex

Clues to the initial stages of complex formation may come from the recent active state structures of rhodopsin^{32,33}. Figure 6a, b compares the active-state structure of β_2 AR in the β_2 AR–Gs complex with the recent structure of metarhodopsin II bound to a peptide representing the carboxy terminus of transducin³². The conformational changes in TM5 and TM6 are smaller in metarhodopsin II, and the position of the carboxy-terminal α helix of transducin is tilted by approximately 30° relative to the position of the homologous region of Gs. These may represent fundamental differences in the receptor–G protein interactions between these two proteins, but given the conservation of the G-protein binding pocket, the changes more probably reflect the more extensive contacts formed with the intact G protein. The position of the transducin peptide in metarhodopsin II may represent the initial interaction between a GDP-bound G protein and a GPCR. We have attempted to reproduce a similar complex between the β_2 AR and a synthetic peptide representing the carboxy-terminal 20 amino acids of Gs, but did not observe any effect of this peptide on receptor function, possibly due to the solubility and behaviour of the peptide in solution. However, when the carboxy-terminal 20 amino acids of Gs are fused to the carboxy terminus of the β_2 AR (Fig. 6c), we observe a 27-fold increase in agonist affinity (Fig. 6d). This effect is only 3.5-fold smaller than the effect we observe on agonist binding affinity in the β_2 AR–Gs complex, and demonstrates that there is a functional interaction between the peptide and receptor that may represent an initial stage in β_2 AR–Gs complex formation. Figure 6e and f presents a possible sequence of interactions of β_2 AR and Gs when forming the nucleotide-free complex. The first interaction of the β_2 AR with the Gs heterotrimer would require a movement of the carboxy terminus of the α 5-helix away from the β 6-strand to permit interactions with the β_2 AR similar to those observed in metarhodopsin II (Fig. 6e). The availability of the carboxy terminus of the α 5-helix for interactions with the β_2 AR is supported by deuterium exchange studies³⁵ showing that this segment is more dynamic in the Gs–GDP heterotrimer than

would be expected from the crystal structure of G α s²⁶. The subsequent formation of more extensive interactions between the β_2 AR ICL2 and the amino terminus of G α s requires a rotation of G α sRas relative to the receptor and would be associated with further conformational changes in both β_2 AR and G α sRas (Fig. 6f). We cannot say when GDP is released during the formation of the complex; however, we speculate that uncoupling of the G α sAH from G α sRas is a consequence of nucleotide release or at least a coincident event. This binding model is in agreement with deuterium exchange experiments³⁵.

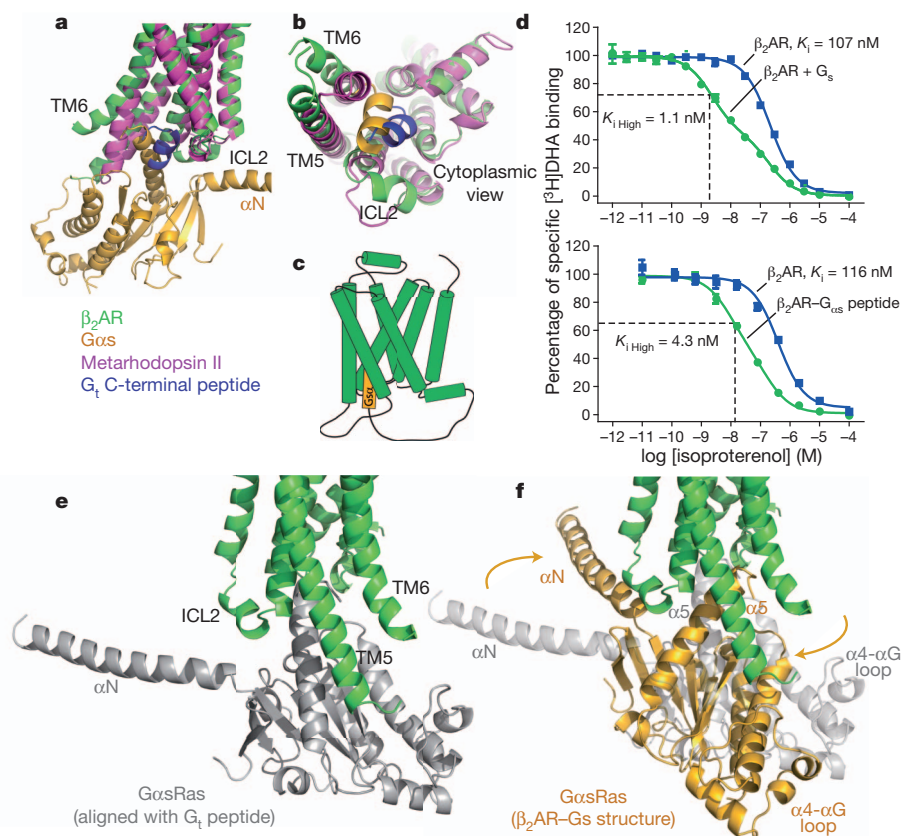
The β_2 AR–Gs complex crystal structure provides the first high-resolution view of transmembrane signalling for a GPCR. We now have a framework to design experiments to investigate the mechanism of complex formation, GTP binding and complex dissociation. Of particular interest will be studies designed to determine the functional significance of the large movement of G α sAH relative to G α sRas that is observed in the β_2 AR–Gs complex. A better understanding of the structural basis for G protein activation may provide new approaches for drug discovery. The high degree of structural homology within the ligand-binding pocket has posed challenges for developing highly selective drugs for specific GPCR targets. In contrast, there is relatively low homology at the interface between the β_2 AR and G α s, so identifying sequence and structural features that define specificity for particular G proteins may enable the development of selective inhibitors of specific GPCR–G protein interactions.

METHODS SUMMARY

The β_2 AR–Gs complex was crystallized from β_2 AR and Gs protein expressed in insect cells. Crystallogensis was aided by fusing T4 lysozyme to the amino terminus of the β_2 AR and the addition of a nanobody (Nb35) that binds at the interface between the G α and G β subunits. Crystals were grown in a LCP using 7.7 MAG, a lipid that accommodates membrane proteins with larger hydrophilic surfaces¹⁵. Diffraction data were measured at beamline 23ID-B of the Advanced Photon Source and the structure was solved by molecular replacement. For more experimental details see Methods.

Figure 6 | Possible sequence of β_2 AR–Gs complex formation.

a, b, Comparison of the β_2 AR–Gs structure (green and orange) with metarhodopsin II³² (PDB ID: 3PQR) (purple) bound with the carboxy-terminal peptide of transducin (G_t) (blue). TM7 has been omitted in panel **a** to better visualize the G proteins. **c**, Cartoon of the β_2 AR–Gs peptide fusion construct used in the binding experiments (**d**). **d**, Competition binding experiments between [³H]dihydroalprenolol ([³H]DHA) and full agonist isoproterenol. Top panel shows binding data (reproduced from ref. 12) on β_2 AR reconstituted in HDL particles with and without Gs heterotrimer. The fraction of β_2 AR in the $K_{i\text{ High}}$ state for the β_2 AR with Gs is 0.55. Bottom panel shows binding to β_2 AR and a β_2 AR–Gs peptide fusion expressed in Sf9 cell membranes. The fraction of β_2 AR in the $K_{i\text{ High}}$ state for the β_2 AR–Gs peptide fusion is 0.68. **e**, The initial interaction of agonist-bound β_2 AR and G α sRas may involve an orientation of the carboxy terminus of G α sRas similar to that of the carboxy-terminal peptide of transducin in the structure of metarhodopsin II. **f**, The final position of G α sRas on the β_2 AR as observed in the β_2 AR–Gs complex.



Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 8 June; accepted 11 July 2011.

Published online 19 July 2011.

- Dixon, R. A. *et al.* Cloning of the gene and cDNA for mammalian β -adrenergic receptor and homology with rhodopsin. *Nature* **321**, 75–79 (1986).
- Rasmussen, S. G. *et al.* Crystal structure of the human β_2 adrenergic G-protein-coupled receptor. *Nature* **450**, 383–387 (2007).
- Rosenbaum, D. M. *et al.* GPCR engineering yields high-resolution structural insights into β_2 -adrenergic receptor function. *Science* **318**, 1266–1273 (2007).
- Lefkowitz, R. J. Seven transmembrane receptors: something old, something new. *Acta Physiol. (Oxf.)* **190**, 9–19 (2007).
- Brandt, D. R., Asano, T., Pedersen, S. E. & Ross, E. M. Reconstitution of catecholamine-stimulated guanosinetriphosphatase activity. *Biochemistry* **22**, 4357–4362 (1983).
- Cerione, R. A. *et al.* The mammalian β_2 -adrenergic receptor: reconstitution of functional interactions between pure receptor and pure stimulatory nucleotide binding protein of the adenylate cyclase system. *Biochemistry* **23**, 4519–4525 (1984).
- Ross, E. M., Maguire, M. E., Sturgill, T. W., Biltonen, R. L. & Gilman, A. G. Relationship between the β -adrenergic receptor and adenylate cyclase. Studies of ligand binding and enzyme activity in purified membranes of S49 lymphoma cells. *J. Biol. Chem.* **252**, 5761–5775 (1977).
- De Lean, A., Stadel, J. M. & Lefkowitz, R. J. A ternary complex model explains the agonist-specific binding properties of the adenylate cyclase-coupled β -adrenergic receptor. *J. Biol. Chem.* **255**, 7108–7117 (1980).
- Fredriksson, R., Lagerström, M. C., Lundin, L. G. & Schiöth, H. B. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol. Pharmacol.* **63**, 1256–1272 (2003).
- Azzi, M. *et al.* β -arrestin-mediated activation of MAPK by inverse agonists reveals distinct active conformations for G protein-coupled receptors. *Proc. Natl Acad. Sci. USA* **100**, 11406–11411 (2003).
- Lefkowitz, R. J. & Shenoy, S. K. Transduction of receptor signals by β -arrestins. *Science* **308**, 512–517 (2005).
- Rasmussen, S. G. *et al.* Structure of a nanobody-stabilized active state of the β_2 adrenoceptor. *Nature* **469**, 175–180 (2011).
- Chae, P. S. *et al.* Maltose-neopentyl glycol (MNG) amphiphiles for solubilization, stabilization and crystallization of membrane proteins. *Nature Methods* **7**, 1003–1008 (2010).
- Sprang, S. R. G protein mechanisms: insights from structural analysis. *Annu. Rev. Biochem.* **66**, 639–678 (1997).
- Misquitta, L. V. *et al.* Membrane protein crystallization in lipidic mesophases with tailored bilayers. *Structure* **12**, 2113–2124 (2004).
- Domanska, K. *et al.* Atomic structure of a nanobody-trapped domain-swapped dimer of an amyloidogenic β_2 -microglobulin variant. *Proc. Natl Acad. Sci. USA* **108**, 1314–1319 (2011).
- Warne, T. *et al.* Structure of a β_1 -adrenergic G-protein-coupled receptor. *Nature* **454**, 486–491 (2008).
- Hofmann, K. P. *et al.* A G protein-coupled receptor at work: the rhodopsin model. *Trends Biochem. Sci.* **34**, 540–552 (2009).
- Moro, O., Lameh, J., Hogger, P. & Sadée, W. Hydrophobic amino acid in the i2 loop plays a key role in receptor-G protein coupling. *J. Biol. Chem.* **268**, 22273–22276 (1993).
- Baltensperger, K. *et al.* The β -adrenergic receptor is a substrate for the insulin receptor tyrosine kinase. *J. Biol. Chem.* **271**, 1061–1064 (1996).
- Jastrzebska, B., Tsybovsky, Y. & Palczewski, K. Complexes between photoactivated rhodopsin and transducin: progress and questions. *Biochem. J.* **428**, 1–10 (2010).
- Johnston, C. A. & Siderovski, D. P. Receptor-mediated activation of heterotrimeric G-proteins: current structural insights. *Mol. Pharmacol.* **72**, 219–230 (2007).
- Breitwieser, G. E. G protein-coupled receptor oligomerization: implications for G protein activation and cell signaling. *Circ. Res.* **94**, 17–27 (2004).
- Bañeres, J. L. & Parello, J. Structure-based analysis of GPCR function: evidence for a novel pentameric assembly between the dimeric leukotriene B4 receptor BLT1 and the G-protein. *J. Mol. Biol.* **329**, 815–829 (2003).
- Angers, S., Salahpour, A. & Bouvier, M. Dimerization: an emerging concept for G protein-coupled receptor ontogeny and function. *Annu. Rev. Pharmacol. Toxicol.* **42**, 409–435 (2002).
- Sunahara, R. K., Tesmer, J. J., Gilman, A. G. & Sprang, S. R. Crystal structure of the adenylyl cyclase activator G_{sz} . *Science* **278**, 1943–1947 (1997).
- Van Eps, N. *et al.* Interaction of a G protein with an activated receptor opens the interdomain interface in the α subunit. *Proc. Natl Acad. Sci. USA* **108**, 9420–9424 (2011).
- Markby, D. W., Onrust, R. & Bourne, H. R. Separate GTP binding and GTPase activating domains of a $G\alpha$ subunit. *Science* **262**, 1895–1901 (1993).
- Conklin, B. R. & Bourne, H. R. Structural elements of $G\alpha$ subunits that interact with $G\beta\gamma$ receptors, and effectors. *Cell* **73**, 631–641 (1993).
- Oldham, W. M. & Hamm, H. E. Heterotrimeric G protein activation by G-protein-coupled receptors. *Nature Rev. Mol. Cell Biol.* **9**, 60–71 (2008).
- Wall, M. A. *et al.* The structure of the G protein heterotrimer $G_{i1}\beta_1\gamma_2$. *Cell* **83**, 1047–1058 (1995).
- Choe, H. W. *et al.* Crystal structure of metarhodopsin II. *Nature* **471**, 651–655 (2011).
- Standfuss, J. *et al.* The structural basis of agonist-induced activation in constitutively active rhodopsin. *Nature* **471**, 656–660 (2011).
- Westfield, G. *et al.* Structural flexibility of the $G\alpha s$ α -helical domain in the β_2 -adrenoceptor G_s complex. *Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.1113645108 (in the press).
- Chung, K. Y. *et al.* β_2 adrenergic receptor-induced conformational changes in the heterotrimeric G protein G_s . *Nature* doi:10.1038/nature10488 (this issue).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge support from National Institutes of Health Grants NS028471 (B.K.K.) and GM083118 (B.K.K. and R.K.S.), GM56169 (W.I.W.), P01 GM75913 (S.H.G.), T32-GM008270 and P60DK-20572 (R.K.S.), GM75915, P50GM073210 and U54GM094599 (M.C.), Science Foundation Ireland (07/IN.1/B1836) and FP7 COST Action CM0902 (M.C.), the Mathers Foundation (B.K.K. and W.I.W.), the Lundbeck Foundation (Junior Group Leader Fellowship, S.G.F.R.), the University of Michigan Biomedical Sciences Scholars Program (R.K.S.), the Fund for Scientific Research of Flanders (FWO-Vlaanderen) and the Institute for the Encouragement of Scientific Research and Innovation of Brussels (ISIRI) (E.P. and J.S.), The Danish Council for Independent Research, Medical Sciences (J.M.M.). We thank D. Grayson and A. Coughlan for help with lipid synthesis.

Author Contributions S.G.F.R. performed the final stages of β_2 AR purification; assisted with β_2 AR and G_s protein virus production and expression in insect cell cultures; worked out conditions to form and stabilize the β_2 AR- G_s complex following screening, identification and characterization of the BL-167107 agonist and MNG-3 detergent; developed the β_2 AR- G_s complex purification strategy with B.K.K. and characterized the stability of the complex under a variety of conditions; purified and analysed all preparations of the β_2 AR- G_s complex used for crystallography, DXMS and electron microscopy studies, immunization, and nanobody selection; expressed and purified nanobodies and characterized β_2 AR- G_s -Nb binding by size exclusion chromatography; set up crystallization trials in detergent solution, bicelles and lipidic cubic phase; crystallized the T4L- β_2 AR- G_s , T4L- β_2 AR- G_s -Nb37 and T4L- β_2 AR- G_s -Nb35 complexes; optimized crystallization conditions and grew crystals for data collection; assisted with data collection and manuscript preparation. B.T.D. managed G_s heterotrimer subunit virus production and titration; expressed and purified G_s protein; with R.K.S. he identified the use of apyrase in forming the β_2 AR- G_s complex and foscarnet/pyrophosphate during crystallogenesis; reconstituted the β_2 AR- G_s complex and receptor alone in high density lipoprotein particles which were used for the initial nanobody selection. He assisted with data collection. Y.Z. designed, generated and optimized the T4L- β_2 AR fusion protein, characterized its expression and functional properties, and prepared baculovirus for large scale expression. A.C.K. collected crystals, collected and processed diffraction data, solved and refined the structure, and assisted with manuscript preparation. K.Y.C. developed the cross-linking conditions for the purified β_2 AR- G_s complex used for immunization of llamas. E.P. performed llama immunization, cloned and expressed nanobodies and performed initial selections. J.S. supervised nanobody production. D.C. assisted with G_s heterotrimer expression and purification. J.M.M. generated the β_2 AR- $G_s\alpha$ peptide fusion construct, expressed it in insect cell membranes and performed competition binding experiments. F.S.T. expressed β_2 AR in insect cell cultures and with T.S.K. performed the initial stage of β_2 AR purification. S.T.A.S., J.A.L. and M.C. provided the 7.7 MAG lipid and helpful suggestions for lipidic mesophase crystallization using this lipid. P.S.C. and S.H.G. provided MNG-3 detergent for stabilization of the β_2 AR- G_s complex. G.S. provided the essential feedback from the electron microscopy studies that helped guide the crystallization effort. W.I.W. oversaw data processing, structure determination and refinement. R.K.S. supervised G_s protein production, provided valuable ideas and insights into G_s structure and function, and assisted with data collection and manuscript preparation. B.K.K. was responsible for overall project strategy and management, harvested crystals and assisted with collection of diffraction data, and wrote the manuscript.

Author Information Coordinates and structure factors for the β_2 AR- G_s complex are deposited in the Protein Data Bank (accession code 3SN6). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to B.K.K. (kobilka@stanford.edu) or R.K.S. (sunahara@umich.edu).

METHODS

Expression and purification of β_2 AR, Gs heterotrimer and nanobody-35. An amino-terminally fused T4 lysozyme- β_2 AR construct with β_2 AR truncated in position 365 (T4L- β_2 AR, described in detail below) was expressed in Sf9 insect cell cultures infected with recombinant baculovirus (BestBac, Expression Systems), and solubilized in *n*-dodecyl- β -D-maltoside (DDM) according to methods described previously³⁶ (see Supplementary Fig. 7 for purification overview). A β_2 AR construct truncated after residue 365 (β_2 AR-365) was used for the majority of the analytical experiments. M1 Flag affinity chromatography (Sigma) served as the initial purification step followed by alprenolol-Sepharose chromatography for selection of functional receptor. A subsequent M1 Flag affinity chromatography step was used to exchange receptor-bound alprenolol for high-affinity agonist BI-167107. The agonist-bound receptor was eluted, dialysed against buffer (20 mM HEPES, pH 7.5, 100 mM NaCl, 0.1% DDM and 10 μ M BI-167107), treated with lambda phosphatase (New England Biolabs), and concentrated to approximately 50 mg ml⁻¹ with a 50 kDa molecular weight cut off (MWCO) Millipore concentrator. Prior to spin concentration, the β_2 AR-365 construct, but not T4L- β_2 AR, was treated with PNGaseF (New England Biolabs) to remove amino-terminal amino-linked glycosylation. The purified receptor was routinely analysed by SDS-PAGE/Coomassie brilliant blue staining (see Supplementary Fig. 8a).

Bovine G α_s short, His₆-rat G β_1 and bovine G γ_2 were expressed in HighFive insect cells (Invitrogen) grown in Insect Xpress serum-free media (Lonza). Cultures were grown to a density of 1.5 million cells per ml and then infected with three separate *Autographa californica* nuclear polyhedrosis viruses each containing the gene for one of the G protein subunits at a 1:1 multiplicity of infection (the viruses were a gift from A. Gilman). After 40–48 h of incubation the infected cells were harvested by centrifugation and resuspended in 75 ml lysis buffer (50 mM HEPES, pH 8.0, 65 mM NaCl, 1.1 mM MgCl₂, 1 mM EDTA, 1 \times PTT (35 μ g ml⁻¹ phenylmethanesulphonyl fluoride, 32 μ g ml⁻¹ tosyl phenylalanyl chloromethyl ketone, 32 μ g ml⁻¹ tosyl lysyl chloromethyl ketone), 1 \times LS (3.2 μ g ml⁻¹ leupeptin and 3.2 μ g ml⁻¹ soybean trypsin inhibitor), 5 mM β -mercaptoethanol (β -ME), and 10 μ M GDP) per litre of culture volume. The suspension was pressurized with 600 p.s.i. N₂ for 40 min in a nitrogen cavitation bomb (Parr Instrument Company). After depressurization, the lysate was centrifuged to remove nuclei and unlysed cells, and then ultracentrifuged at 180,000g for 40 min. The pelleted membranes were resuspended in 30 ml wash buffer (50 mM HEPES, pH 8.0, 50 mM NaCl, 100 μ M MgCl₂, 1 \times PTT, 1 \times LS, 5 mM β -ME, 10 μ M GDP) per litre culture volume using a Dounce homogenizer and centrifuged again at 180,000g for 40 min. The washed pellet was resuspended in a minimal volume of wash buffer and flash-frozen with liquid nitrogen.

The frozen membranes were thawed and diluted to a total protein concentration of 5 mg ml⁻¹ with fresh wash buffer. Sodium cholate detergent was added to the suspension at a final concentration of 1.0%, MgCl₂ was added to a final concentration of 5 mM, and 0.05 mg of purified protein phosphatase 5 (prepared in house) was added per litre of culture volume. The sample was stirred on ice for 40 min, and then centrifuged at 180,000g for 40 min to remove insoluble debris. The supernatant was diluted fivefold with Ni-NTA load buffer (20 mM HEPES, pH 8.0, 363 mM NaCl, 1.25 mM MgCl₂, 6.25 mM imidazole, 0.2% Anzergent 3-12, 1 \times PTT, 1 \times LS, 5 mM β -ME, 10 μ M GDP), taking care to add the buffer slowly to avoid dropping the cholate concentration below its critical micelle concentration too quickly. Ni-NTA resin (3 ml; Qiagen) pre-equilibrated in Ni-NTA wash buffer 1 (20 mM HEPES, pH 8.0, 300 mM NaCl, 2 mM MgCl₂, 5 mM imidazole, 0.2% cholate, 0.15% Anzergent 3-12, 1 \times PTT, 1 \times LS, 5 mM β -ME, 10 μ M GDP) per litre culture volume was added and the sample was stirred on ice for 20 min. The resin was collected into a gravity column and washed with 4 \times column volumes of Ni-NTA wash buffer 1, Ni-NTA wash buffer 2 (20 mM HEPES, pH 8.0, 50 mM NaCl, 1 mM MgCl₂, 10 mM imidazole, 0.15% Anzergent 3-12, 0.1% DDM, 1 \times PTT, 1 \times LS, 5 mM β -ME, 10 μ M GDP), and Ni-NTA wash buffer 3 (20 mM HEPES, pH 8.0, 50 mM NaCl, 1 mM MgCl₂, 5 mM imidazole, 0.1% DDM, 1 \times PTT, 1 \times LS, 5 mM β -ME, 10 μ M GDP). The protein was eluted with Ni-NTA elution buffer (20 mM HEPES, pH 8.0, 40 mM NaCl, 1 mM MgCl₂, 200 mM imidazole, 0.1% DDM, 1 \times PTT, 1 \times LS, 5 mM β -ME, 10 μ M GDP). Protein-containing fractions were pooled and MnCl₂ was added to a final concentration of 100 μ M. Purified lambda protein phosphatase (50 μ g; prepared in house) was added per litre of culture volume and the eluate was incubated on ice with stirring for 30 min. The eluate was passed through a 0.22- μ m filter and loaded directly onto a MonoQ HR 16/10 column (GE Healthcare) equilibrated in MonoQ buffer A (20 mM HEPES, pH 8.0, 50 mM NaCl, 100 μ M MgCl₂, 0.1% DDM, 5 mM β -ME, 1 \times PTT). The column was washed with 150 ml buffer A at 5 ml min⁻¹ and bound proteins were eluted over 350 ml with a linear gradient up to 28% MonoQ buffer B (same as buffer A except with 1 M NaCl). Fractions were collected in tubes spotted with enough GDP to make a final concentration of

10 μ M. The Gs-containing fractions were concentrated to 2 ml using a stirred ultrafiltration cell (Amicon) with a 10-kDa nominal molecular weight limit (NMWL) regenerated cellulose membrane (Millipore). The concentrated sample was run on a Superdex 200 prep grade XK 16/70 column (GE Healthcare) equilibrated in S200 buffer (20 mM HEPES, pH 8.0, 100 mM NaCl, 1.1 mM MgCl₂, 1 mM EDTA, 0.012% DDM, 100 μ M TCEP, 2 μ M GDP). The fractions containing pure Gs were pooled, glycerol was added to 10% final concentration, and then the protein was concentrated to at least 10 mg ml⁻¹ using a 30 kDa MWCO centrifugal ultrafiltration device (Millipore). The concentrated sample was then aliquoted, flash frozen, and stored at -80 °C. A typical yield of final, purified Gs heterotrimer from 8 l of cell culture volume was 6 mg.

Nanobody-35 (Nb35) was expressed in the periplasm of *E. coli* strain WK6, extracted, and purified by nickel affinity chromatography according to previously described methods¹² followed by ion-exchange chromatography (Supplementary Fig. 9a) using a Mono S 10/100 GL column (GE Healthcare). Selected Nb35 fractions were dialysed against buffer (10 mM HEPES, pH 7.5, 100 mM NaCl) and concentrated to approximately 65 mg ml⁻¹ with a 10 kDa MWCO Millipore concentrator.

Complex formation, stabilization and purification. Formation of a stable complex (see Supplementary Fig. 10) was accomplished by mixing Gs heterotrimer at approximately 100 μ M concentration with BI-167107-bound T4L- β_2 AR (or β_2 AR-365) in molar excess (approximately 130 μ M) in 2 ml buffer (10 mM HEPES, pH 7.5, 100 mM NaCl, 0.1% DDM, 1 mM EDTA, 3 mM MgCl₂, 10 μ M BI-167107) and incubating for 3 h at room temperature. BI-167107, which was identified from screening and characterizing approximately 50 different β_2 AR agonists (data not shown), has a dissociation half-time of approximately 30 h, providing a higher degree of stabilization to the active G protein-bound receptor than other full agonists such as isoproterenol¹². To maintain the high-affinity nucleotide-free state of the complex, apyrase (25 mU ml⁻¹, NEB) was added after 90 min to hydrolyse residual GDP released from G α_s upon binding to the receptor. GMP resulting from hydrolysis of GDP by apyrase has very poor affinity for the G protein in the complex. Rebinding of GDP can cause dissociation of the β_2 AR-Gs complex (Supplementary Fig. 1a).

The β_2 AR-Gs complex in DDM shows significant dissociation after 48 h at 4 °C (Supplementary Fig. 11a). We screened and characterized over 50 amphiphiles (data not shown) and identified MNG-3 (refs 12, 13; NG-310, Affymetrix-Anatrace) and its closely related analogues as detergents that substantially stabilize the complex (Supplementary Fig. 11a). The complex was exchanged into MNG-3 by adding the β_2 AR-Gs mixture (2 ml) to 8 ml buffer (20 mM HEPES, pH 7.5, 100 mM NaCl, 10 μ M BI-167107) containing 1% MNG-3 for 1 h at room temperature.

At this stage the mixture contains the β_2 AR-Gs complex, non-functional Gs, and an excess of β_2 AR. To separate functional β_2 AR-Gs complex from non-functional Gs, and to complete the detergent exchange, the β_2 AR-Gs complex was immobilized on M1 Flag resin and washed in buffer (20 mM HEPES, pH 7.5, 100 mM NaCl, 10 μ M BI-167107, and 3 mM CaCl₂) containing 0.2% MNG-3. To prevent cysteine bridge-mediated aggregation of β_2 AR-Gs complexes, 100 μ M TCEP was added to the eluted protein before concentrating it with a 50 kDa MWCO Millipore concentrator. Of note, it was discovered later that crystal growth improved at even higher TCEP concentrations (above 1 mM) compared to 100 μ M TCEP, and that the integrity of the β_2 AR-Gs complex in MNG-3 was stable to 10 mM TCEP as measured by gel filtration analysis (Supplementary Fig. 12). In contrast, DDM-solubilized β_2 AR loses its ability to bind the high-affinity antagonist [³H]dihydroalprenolol ([³H]DHA) in 10 mM TCEP (data not shown), probably due to disruption of extracellular disulphide bonds. Iodoacetamide could not be used to block reactive cysteines on Gs α and β subunits as it caused dissociation of the β_2 AR-Gs complex (Supplementary Fig. 12b). The final size exclusion chromatography procedure to separate excess free receptor from the β_2 AR-Gs complex (Supplementary Fig. 8b) was performed on a Superdex 200 10/300 GL column (GE Healthcare) equilibrated with buffer containing 0.02% MNG-3, 10 mM HEPES, pH 7.5, 100 mM NaCl, 10 μ M BI-167107 and 100 μ M TCEP. Peak fractions were pooled (Supplementary Fig. 8b) and concentrated to approximately 90 mg ml⁻¹ with a 100 kDa MWCO Viva-spin concentrator and analysed by SDS-PAGE/Coomassie brilliant blue staining (Supplementary Fig. 8a) and gel filtration (Supplementary Fig. 8c). To confirm a pure, homogeneous and dephosphorylated preparation, the β_2 AR-Gs complex was routinely analysed by ion exchange chromatography (Supplementary Fig. 8d).

Protein engineering. To increase the probability of obtaining crystals of the β_2 AR-Gs complex we set out to increase the polar surface area on the extracellular side of the receptor using two strategies. The first approach, to generate extracellular binding antibodies, was not successful. The second approach was to replace the flexible and presumably unstructured amino terminus with the globular protein T4 lysozyme (T4L) used previously to crystallize and solve the carazolol-bound receptor³. The construct used here (T4L- β_2 AR) contained the

cleavable signal sequence followed by the M1 Flag epitope (DYKDDDDA), the Tobacco etch virus (TEV) protease recognition sequence (ENLYFQG), bacteriophage T4 lysozyme from N2 through Y161 including C54T and C97A mutations, and a two residue alanine linker fused to the human β_2 AR sequence D29 through G365. The PNGaseF-inaccessible glycosylation site of the β_2 AR at N187 was mutated to Glu. M96 and M98 in the first extracellular loop were each replaced by Thr to increase the otherwise low expression level of T4L- β_2 AR. The threonine mutations did not affect ligand binding affinity for [3 H]DHA, but caused a small, approximately twofold decrease in affinity for isoproterenol (data not shown).

The β_2 AR-Gs peptide fusion construct used for [3 H]DHA competition binding with isoproterenol was constructed from the receptor truncated at position 365 and fused to the last 21 amino acids of the G α s subunit (amino acids 374–394, except for C379A). A Gly-Ser is inserted between the receptor and the peptide. Also an extended TEV protease site (SENLYFQGS) was introduced in the β_2 AR between G360 and G361.

Stabilization of Gs with nanobodies. From negative stain electron microscopy imaging³⁴, we observed that the α -helical domain of G α s was flexible and therefore possibly responsible for poor crystal quality. Targeted stabilization of this domain was addressed by immunizing two llamas (*Lama glama*) with the bis(sulphosuccinimidyl)glutarate (BS2G, Pierce) cross-linked β_2 AR-Gs-BI-167107 ternary complex. Peripheral blood lymphocytes were isolated from the immunized animals to extract total RNA, prepare cDNA and construct a Nanobody phage display library according to published methods¹⁶. Nb35 and Nb37 were enriched by two rounds of biopanning on the β_2 AR-Gs-BI-167107 ternary complex embedded in biotinylated high-density lipoprotein particles³⁷. Nb35 and Nb37 were selected for further characterization because they bind the β_2 AR-Gs-BI-167107 ternary complex but not the free receptor in an ELISA assay. Nanobody binding to the β_2 AR-Gs complex was confirmed by size exclusion chromatography (Supplementary Fig. 1d), and it was noted that both nanobodies protected the complex from dissociation by GTP γ S, suggestive of a stabilizing Gs-Nb interaction (Supplementary Fig. 1d).

Crystallization. BI-167107 bound T4L- β_2 AR-Gs complex and Nb35 were mixed in 1:1.2 molar ratio. The small molar excess of Nb35 was verified by analytical gel filtration (see Supplementary Fig. 9b). The mixture incubated for 1 h at room temperature before mixing with 7.7 MAG (provided by M. Caffrey) containing 10% cholesterol (C8667, Sigma) in 1:1 protein solution to lipid ratio (w/w) using the twin-syringe mixing method reported previously³⁸. The concentration of T4L- β_2 AR-Gs-Nb35 complex in 7.7 MAG was approximately 25 mg ml⁻¹. We believe the detergent MNG-3 stabilizes the T4L- β_2 AR-Gs complex during its incorporation into the lipid cubic phase. This may be due to the high affinity of MNG-3 for the receptor. The β_2 AR in MNG-3 maintains its structural integrity even when diluted below the CMC of the detergent, in contrast to β_2 AR in DDM, which rapidly loses binding activity (Supplementary Fig. 11b). Moreover, MNG-3 improved crystal size and quality, as previously reported^{12,13,39}. The protein:lipid mixture was delivered through an LCP dispensing robot (Gryphon, Art Robbins Instruments) in 40 nl drops to either 24-well or 96-well glass sandwich plates and overlaid en-bloc with 0.8 μ l precipitant solution. Multiple crystallization leads were initially identified using in-house screens partly based on reagents from the StockOptions Salt kit (Hampton Research). Crystals for data collection were grown in 18 to 22% PEG 400, 100 mM MES, pH 6.5 (Supplementary Fig. 1c), 350 to 450 mM potassium nitrate, 10 mM foscarnet (Supplementary Fig. 1b), 1 mM TCEP (Supplementary Fig. 12c), and 10 μ M BI-167107. Crystals reached full size within 3–4 days at 20 °C and were picked from a sponge-like mesophase and flash-frozen in liquid nitrogen without additional cryoprotectant.

Microcrystallography data collection and processing. Diffraction data were measured at the Advanced Photon Source beamline 23 ID-B. Hundreds of crystals were screened, and a final data set was compiled using diffraction wedges of typically 10 degrees from 20 strongly diffracting crystals (Supplementary Table 2). All data reduction was performed using HKL2000 (ref. 40). Although in many cases diffraction to beyond 3 Å was seen in initial frames, radiation damage and anisotropic diffraction resulted in low completeness in higher resolution shells. Analysis of the final data set by the UCLA diffraction anisotropy server⁴¹ indicated that diffraction along the a^* axis was superior to that in other directions. On

the basis of an $F/\sigma(F)$ cutoff of 3 along each reciprocal space axis, reflections were subjected to an anisotropic truncation with resolution limits of 2.9, 3.2 and 3.2 Å along a^* , b^* and c^* before use in refinement. We report this structure to an overall resolution of 3.2 Å. Despite the low completeness in the highest resolution shells (Supplementary Table 3) inclusion of these reflections gave substantial improvements in map quality and lower R_{free} during refinement.

Structure solution and refinement. The structure was solved by molecular replacement using Phaser^{42,43}. The order of the molecular replacement search was found to be critical in solving the structure. In order, the search models used were: the β and γ subunits from a Gi heterotrimer (PDB ID: 1GP2), the Gs α Ras-like domain (PDB ID: 1AZT), the active-state β_2 AR (PDB ID: 3P0G), a β_2 AR-binding nanobody (PDB ID: 3P0G), T4 lysozyme (PDB ID: 2RH1), and the Gs α -helical domain (PDB ID: 1AZT). Following the determination of the initial structure by molecular replacement, rigid body refinement and simulated annealing were performed in Phenix⁴⁴ and BUSTER⁴⁵, followed by restrained refinement and manual rebuilding in Coot⁴⁶. After iterative refinement and manual adjustments, the structure was refined in CNS using the DEN method⁴⁷. Although the resolution of this structure exceeds that for which DEN is typically most useful, the presence of several poorly resolved regions indicated that the incorporation of additional information to guide refinement could provide better results. The DEN reference models used were those used for molecular replacement, with the exception of NB35, which was well ordered and for which no higher resolution structure is available. Side chains were omitted from 53 residues for which there was no electron density past CB below a low contour level of 0.7 σ in a 2Fo – Fc map. Figures were prepared using PyMOL (The PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LLC.). MolProbity was used to determine Ramachandran statistics⁴⁸.

Competition binding. Membranes expressing the β_2 AR or the β_2 AR-Gs peptide fusion were prepared from baculovirus-infected Sf9 cells and [3 H]DHA-binding performed as previously described⁴⁹. For competition binding, membranes were incubated with [3 H]DHA (1.1 nM final) and increasing concentrations of (–)-isoproterenol for 1 h before harvesting onto GF/B filters. Competition data were fitted to a two-site binding model and isoproterenol high and low K_i values and fractions calculated using GraphPad prism.

36. Kobilka, B. K. Amino and carboxyl terminal modifications to facilitate the production and purification of a G protein-coupled receptor. *Anal. Biochem.* **231**, 269–271 (1995).
37. Whorton, M. R. et al. A monomeric G protein-coupled receptor isolated in a high-density lipoprotein particle efficiently activates its G protein. *Proc. Natl Acad. Sci. USA* **104**, 7682–7687 (2007).
38. Caffrey, M. & Cherezov, V. Crystallizing membrane proteins using lipidic mesophases. *Nature Protocols* **4**, 706–731 (2009).
39. Rosenbaum, D. M. et al. Structure and function of an irreversible agonist- β_2 adrenoceptor complex. *Nature* **469**, 236–240 (2011).
40. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
41. Strong, M. et al. Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA* **103**, 8060–8065 (2006).
42. McCoy, A. J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. D* **63**, 32–41 (2007).
43. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
44. Afonine, P. V., Grosse-Kunstleve, R. W. & Adams, P. D. A robust bulk-solvent correction and anisotropic scaling procedure. *Acta Crystallogr. D* **61**, 850–855 (2005).
45. Blanc, E. et al. Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr. D* **60**, 2210–2221 (2004).
46. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
47. Schröder, G. F., Levitt, M. & Brunger, A. T. Super-resolution biomolecular crystallography with low-resolution data. *Nature* **464**, 1218–1222 (2010).
48. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
49. Swaminath, G., Steenhuis, J., Kobilka, B. & Lee, T. W. Allosteric modulation of β_2 -adrenergic receptor by Zn²⁺. *Mol. Pharmacol.* **61**, 65–72 (2002).

Crystal structure of nucleotide-free dynamin

Katja Faelber¹, York Posor^{2*}, Song Gao^{1,2*}, Martin Held^{3*}, Yvette Roske^{1*}, Dennis Schulze¹, Volker Haucke², Frank Noé³ & Oliver Daumke^{1,4}

Dynamin is a mechanochemical GTPase that oligomerizes around the neck of clathrin-coated pits and catalyses vesicle scission in a GTP-hydrolysis-dependent manner. The molecular details of oligomerization and the mechanism of the mechanochemical coupling are currently unknown. Here we present the crystal structure of human dynamin 1 in the nucleotide-free state with a four-domain architecture comprising the GTPase domain, the bundle signalling element, the stalk and the pleckstrin homology domain. Dynamin 1 oligomerized in the crystals via the stalks, which assemble in a criss-cross fashion. The stalks further interact via conserved surfaces with the pleckstrin homology domain and the bundle signalling element of the neighbouring dynamin molecule. This intricate domain interaction rationalizes a number of disease-related mutations in dynamin 2 and suggests a structural model for the mechanochemical coupling that reconciles previous models of dynamin function.

Dynamin, the founding member of the dynamin superfamily, is a 100-kDa mechanochemical enzyme (Fig. 1a) involved in the scission of clathrin-coated vesicles from the plasma membrane¹. The brain-specific isoform dynamin 1 mediates uptake of synaptic vesicles in presynaptic terminals^{2–4}, whereas a function of dynamin 3 at the post-synaptic density has been described⁵. Dynamin 2 is ubiquitously expressed⁶, and mutations in its middle domain (MD), pleckstrin homology (PH) domain and GTPase effector domain (GED) are linked

to human diseases, for example, rare forms of centronuclear myopathy and Charcot-Marie-Tooth peripheral neuropathy⁷. Upon recruitment via the carboxy-terminal proline-rich domain (PRD), dynamin oligomerizes into helical structures around the neck of budding vesicles and catalyses vesicle scission in a GTP-hydrolysis-dependent manner^{8,9}. Different mechanisms for the scission reaction have been proposed, including GTP-hydrolysis-dependent constriction¹⁰, extension¹¹ and twisting⁹ of the vesicle neck. Other models suggest that the GTP-bound

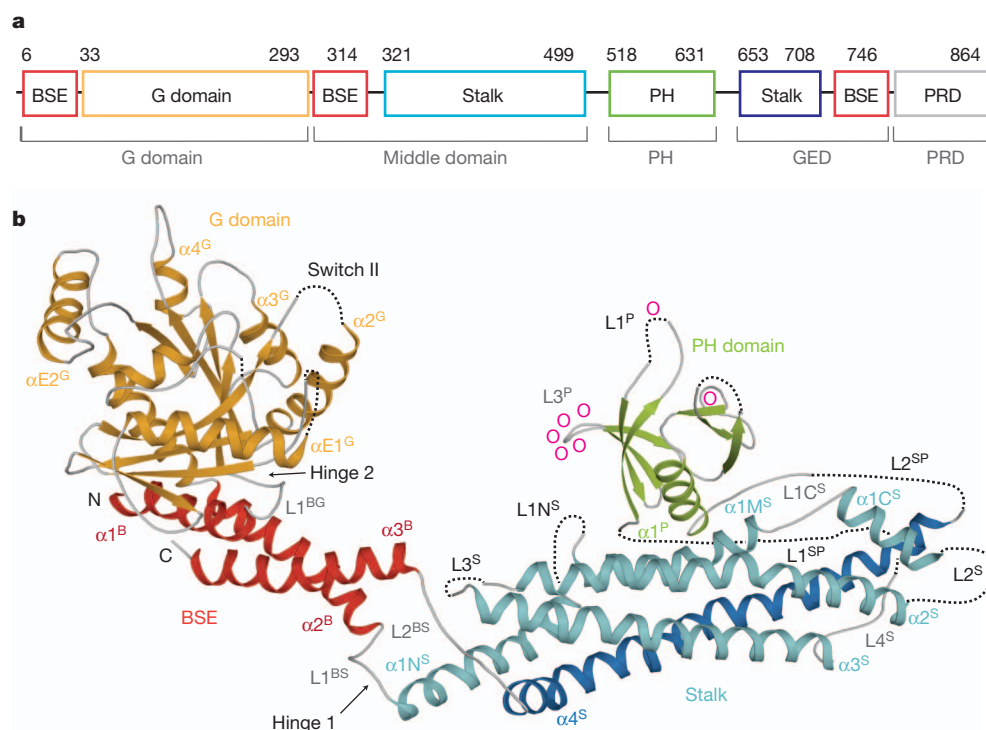


Figure 1 | Structure of nucleotide-free human dynamin 1.

a, Structure-based domain architecture of human dynamin 1. The classical domain assignment is indicated below. **b**, Ribbon-type representation of human dynamin 1. Regions not resolved in the crystal structure are indicated by dotted lines. Domains, distinct secondary structure elements and N and C termini are labelled. Lipid-binding residues are indicated as o.

¹Crystallography, Max-Delbrück-Centrum for Molecular Medicine, Robert-Rössle-Straße 10, 13125 Berlin, Germany. ²Institute for Chemistry and Biochemistry, Freie Universität Berlin, Takustraße 6, 14195 Berlin, Germany. ³Institute for Mathematics, Freie Universität Berlin, Arnimallee 6, 14195 Berlin, Germany. ⁴Institute for Medical Physics and Biophysics, Charité, Ziegelstraße 5-9, 10117 Berlin, Germany. *These authors contributed equally to this work.

dynamain oligomer induces hemifusion of the inner membrane leaflet followed by complete membrane scission after GTP-hydrolysis-dependent release^{12,13}. To resolve the detailed molecular mechanism, high-resolution structural data for full-length dynamain are required which, to date, are available only for the isolated PH domain^{14,15} and the GTPase (G) domain^{16,17}. Low-resolution electron microscopy reconstructions of dynamain oligomers showed that nucleotide binding leads to constriction of helical assemblies through rearrangements in the stalk region composed of the MD and GED¹⁸. Furthermore, G domain dimerization via a conserved interface across the nucleotide-binding site was shown to mediate the stimulated GTPase activity¹⁹. We recently described the structure of the stalk of the dynamain-like antiviral myxovirus resistance protein 1 (MxA) GTPase and elucidated its mode of oligomerization, which involves three distinct interfaces and two loop regions in the stalk²⁰. Using this information, we succeeded to determine the structure of dynamain 1.

The structure of human dynamain 1

We reasoned that the propensity of dynamain to oligomerize at high protein concentrations might interfere with the formation of protein crystals. On the basis of our previous MxA study²⁰, we assayed a number of mutants in a human dynamain 1 construct (amino acids 6–746, Fig. 1a) for oligomerization defects. Indeed, a five-amino-acid exchange (IHGIR395–399AAAAA) in a conserved motif mapping to loop L2 of the MxA stalk²⁰ interfered with higher-order assembly and resulted in a monodisperse dimeric dynamain 1 species (Supplementary Fig. 1, see also ref. 21). Crystals of a construct containing additionally the K562E mutation were obtained in the absence of nucleotides and diffracted to a maximal resolution of 3.7 Å (Supplementary Table 1). The structure was solved by molecular replacement and refined to $R_{\text{work}}/R_{\text{free}}$ of 28.4%/33.5% (Supplementary Table 1). To verify the sequence, the positions of 19 internal methionines were assigned by a single anomalous dispersion approach (Supplementary Fig. 2).

Dynamain 1 (ΔPRD) has a four-domain architecture, composed of the G domain, the bundle signalling element (BSE), the stalk and the PH domain (annotated as superscript G, B, S and P, respectively), which does not strictly follow the sequence-derived domain boundaries (Fig. 1 and

Supplementary Fig. 3). The structure of the amino-terminal G domain is very similar to that of the isolated nucleotide-free G domain¹⁷ (root-mean-square deviation (r.m.s.d.) of 1.4 Å for 287 Cα atoms) and shows a curved central β-sheet surrounded by α-helices at both sides. The two switch regions known to mediate nucleotide-dependent conformational changes and the *cis* stabilizing loop, involved in G domain dimerization¹⁹, are partly disordered. At the N and C termini of the G domain, helices α1^B and α2^B, together with α3^B from the C-terminal part of the GED of the same molecule, form a three-helix bundle, the BSE^{19,22} (Fig. 1 and Supplementary Fig. 4). Compared to the previously described G domain-BSE construct in the GDP•AlF₄[−]-bound form¹⁹, α1^B is shifted by two turns relative to α2^B/α3^B (Supplementary Figs 2 and 4), whereas it interacts with the G domain in a similar fashion in both structures.

At the C-terminal end of α2^B, the BSE connects to the stalk of dynamain 1. The stalk is composed of a four-helix bundle where three helices, α1^S–α3^S, are provided by the MD and α4^S by the N-terminal part of the GED (Fig. 1 and Supplementary Fig. 5a). α1^S in dynamain 1 is subdivided into α1N^S, α1M^S and α1C^S by two disordered loops, L1N^S and L1C^S, compared to a single break of the corresponding helix in MxA. Furthermore, α3^S in dynamain 1 is extended by a highly conserved loop L4^S. At the C terminus of the stalk, α4^S closely packs against α1^S–α3^S via hydrophobic contacts and leads the polypeptide chain back to the BSE. Despite an overall sequence identity of only 16%, the architecture of the dynamain 1 and MxA stalk is remarkably similar (r.m.s.d. of 2.6 Å for 160 aligned Cα atoms, Supplementary Fig. 5).

The PH domain is interconnected between α3^S and α4^S of the stalk by two disordered loops, L1^{SP} and L2^{SP}, and shows only minor deviations from the isolated PH domain of dynamain 1 (refs 14, 15; r.m.s.d. of 0.8 Å for 102 Cα atoms). The three lipid binding loops^{23–25} are only partially resolved and point towards the solvent (Fig. 1b and Supplementary Fig. 3).

Dynamain assembly via the stalk

Similarly to the MxA stalks²⁰, the dynamain 1 stalks in the crystals were arranged in a criss-cross fashion resulting in a linear stalk filament. The highly conserved symmetric interface-2 of 1,200 Å² is located in the centre of the stalk (according to the MxA nomenclature, Fig. 2 and

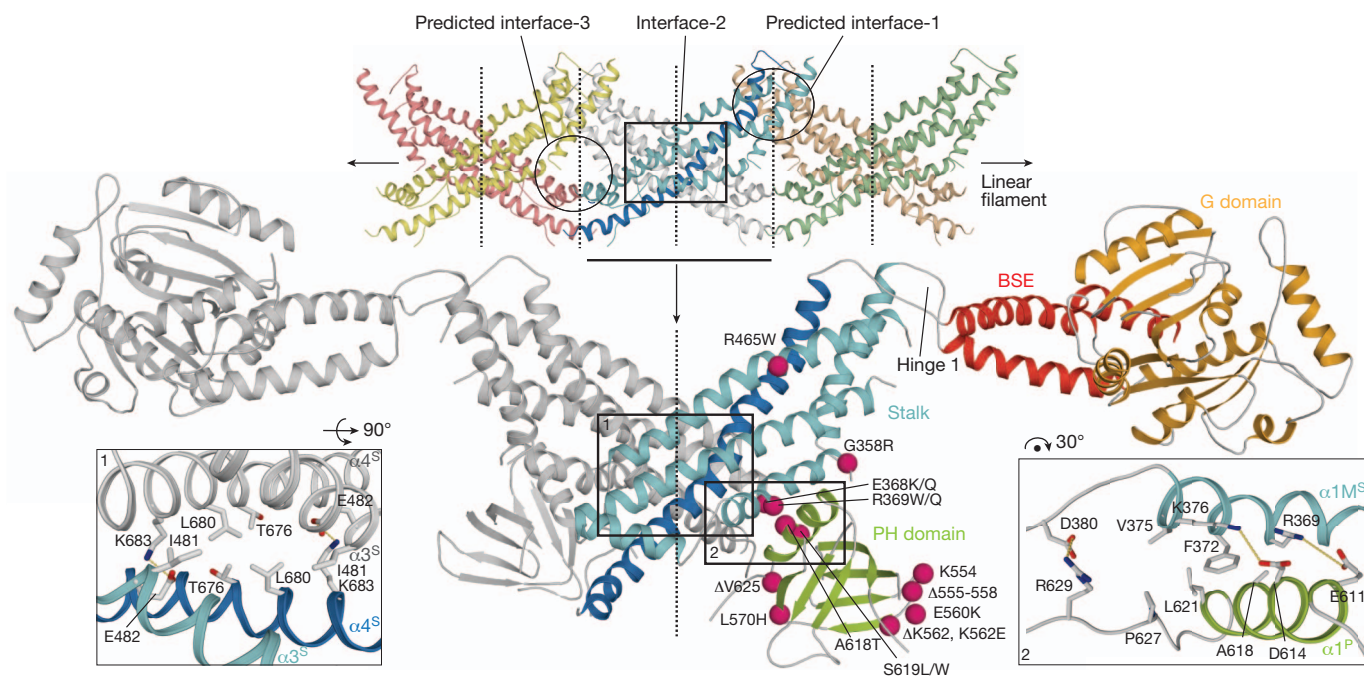


Figure 2 | The dynamain 1 dimer. In the crystals, stalks were arranged in a criss-cross fashion via crystallographic two-fold axis (black dotted lines). Assembly via the central interface-2 leads to an extended dynamain 1 dimer. Black rectangles indicate stalk interfaces shown in detail in the insets (see

Supplementary Fig. 2 for PH domain assignment). Disease-related dynamain 2 mutations causing centronuclear myopathy or Charcot–Marie–Tooth⁷ are represented by pink spheres, with the amino acid exchange indicated.

Supplementary Figs 3 and 5). Assembly via this interface results in an extended dynamin dimer that serves as building block for dynamin oligomers²⁶. Shape and dimension of this dimer agree well with a small-angle X-ray scattering study²⁷.

We previously showed that a second hydrophobic interface in the MxA stalk, interface-1, mediates assembly of higher-order oligomers²⁰. In dynamin 1, however, the stalks do not contact each other directly at the predicted interface-1 (minimal distance 4.5 Å, Supplementary Fig. 5). This difference is caused by a 5° tilt of the dynamin 1 stalks relative to the stalk axis. The hydrophobic nature of this surface in dynamin 1 and its conservation in the dynamin family (Supplementary Fig. 5c) is indicative of a similar function as in MxA as an oligomerization site. Closure of interface-1 might induce a pitch in the dynamin assemblies leading to helical oligomers rather than to ring-like structures as in MxA²⁸.

L2^S containing the disruptive IHGIR395–399AAAAA mutation and L1N^S are not ordered in the linear dynamin 1 oligomer (Supplementary Fig. 5). The corresponding loops in the MxA stalk form a third interface (interface-3) which also contributes to the assembly of oligomers (Fig. 2). Accordingly, mutations in both loops in dynamin 1 (ref. 21), MxA²⁰ and in L1 of dynamin 1-like protein^{29,30} prevent oligomerization. Interestingly, Ser 347 and Tyr 354 in loop L1N^S in dynamin 1 are phosphorylated *in vivo*^{31,32} and might control the assembly status.

Interactions of the BSE

The BSE interacts with the central β -sheet of the G domain via a mostly hydrophilic interface of 1,100 Å² (Supplementary Figs 4 and 6). In contrast, the BSE and concomitantly the G domain are only loosely associated with the stalk of the same molecule via loops L1^{BS} and L2^{BS} constituting a flexible hinge (hinge 1), as observed in other dynamin related proteins³³ (Figs 1b and 2, and Supplementary Fig. 2).

Interestingly, Asp 744 at the C terminus of $\alpha 3^B$ of the BSE contacts Arg 440 in $\alpha 2^S$ of the neighbouring, parallel dynamin 1 stalk (Fig. 3a). A similar intermolecular interaction mediates oligomerization and the antiviral function of MxA (S.G., K.F., O.D., unpublished observation). We tested the importance of this contact experimentally. The wild-type dynamin 1 construct bound efficiently to liposomes resulting in an approximately 200-fold stimulation of GTPase activity (Supplementary Fig. 7). The single R440A and D744A mutants behaved as wild type in these assays. To analyse the role of these residues for clathrin-mediated endocytosis, dynamin 2-eGFP mutants (a fusion of dynamin 2 with enhanced green fluorescent protein) were re-expressed in HeLa cells depleted of endogenous dynamin 2. Both R440A and D744A mutants localized similarly as wild-type dynamin 2 to the plasma membrane (Supplementary Fig. 8), but transferrin internalization was increased (Fig. 3b and Supplementary Fig. 9). Thus, the salt bridge has an inhibitory and/or control function in dynamin-based endocytosis.

The stalk–PH domain interface

$\alpha 1M^S$ of the stalk forms a conserved surface of 370 Å² with the PH domain (Figs 1b and 2, and Supplementary Figs 3 and 5c). Interestingly, 19 unique mutations causing centronuclear myopathy or Charcot–Marie–Tooth disease⁷ cluster in the stalk or the PH domain of dynamin 2, but none localizes to the G domain or BSE (Fig. 2). For example, mutations E368K/Q and R369W/Q in the stalk and A618T and S619L/W in the PH domain are directly in or in close vicinity to the interface between the two domains. Mutations A618T and S619L/W lead to increased oligomerization rates of dynamin in solution^{27,34}, suggesting that the stalk–PH interface controls oligomerization. In gel filtration, the disease-related stalk mutant E368K also eluted as high molecular weight species. Consistently, this mutant showed a 20-fold increased basal GTPase rate, whereas the liposome-stimulated GTPase reaction and transferrin uptake were unchanged (Fig. 3b and Supplementary Fig. 7). In contrast, the R369W mutant behaved as wild-type dynamin in biochemical and endocytosis assays. Mutations at the periphery of

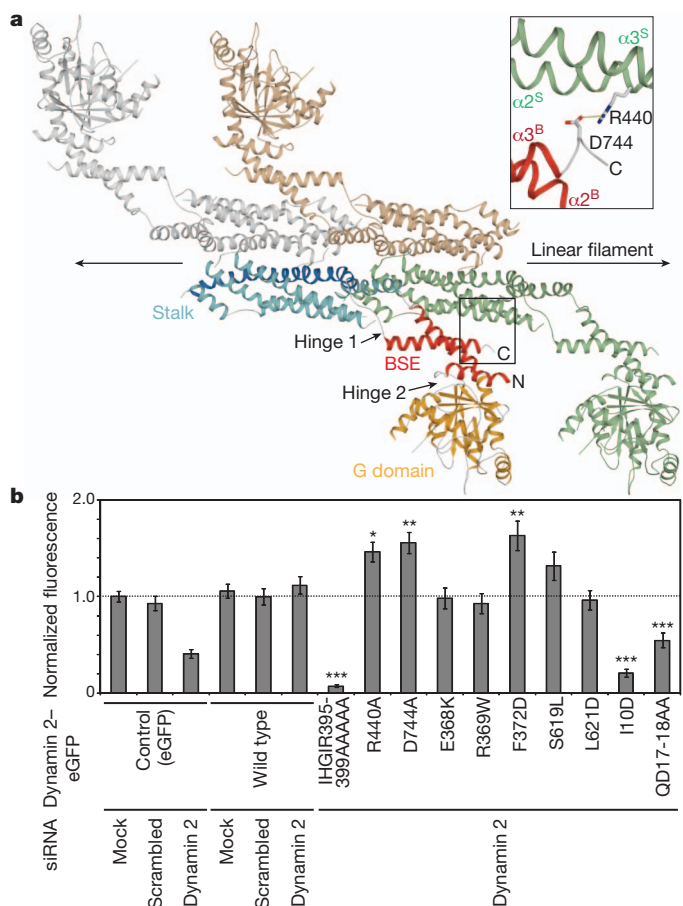


Figure 3 | Stalk interactions with the BSE and PH domain. **a**, Top view on the dynamin 1 oligomer. The PH domains are not drawn for clarity. The insert shows a close view of the intermolecular BSE–stalk interaction. **b**, HeLa cells depleted of endogenous dynamin 2 by short interfering RNA (siRNA) were transfected with a plasmid encoding siRNA-resistant dynamin 2–eGFP and allowed to endocytose fluorescently labelled transferrin. In transfected cells, fluorescence was quantified and normalized to mock-treated cells expressing eGFP. Data shown represent mean \pm standard error; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.0001$ for wild-type versus mutant dynamin 2–eGFP, as determined by *t*-test.

this interface (S619L, L621D) also did not compromise dynamin 2-based endocytic activity (Fig. 3b), indicating that more subtle changes lead to the disease-phenotype. Interestingly, the F372D mutant in the centre of the interface showed significantly higher transferrin uptake compared to wild-type dynamin 2, pointing also to an inhibitory and/or control function of this interface for dynamin-based endocytosis.

Conformational changes during dynamin assembly

Based on the isolated BSE and PH domain (Fig. 1b), the stalk dimer (Fig. 2), the GDP·AlF₄[−]-bound G domain dimer¹⁹, and electron microscopy reconstructions of oligomerized dynamin 1 in the constricted state¹⁸, we generated a molecular model for self-assembly of dynamin into helical structures (Fig. 4a and Supplementary Fig. 10). The resulting dynamin dimer spans a length of more than 260 Å and covers 95° of a dynamin ring, thereby placing G domain and BSE above the neighbouring stalk. According to this model, oligomerization of dynamin proceeds along the central stalk, whereas the G domains mediate contacts between neighbouring turns (see ref. 20 for a similar MxA model).

When comparing the linear structure of dynamin 1 in the crystal with the helical assembly, large-scale domain movements are apparent (Supplementary Fig. 11). The G domain–BSE unit is shifted and rotated

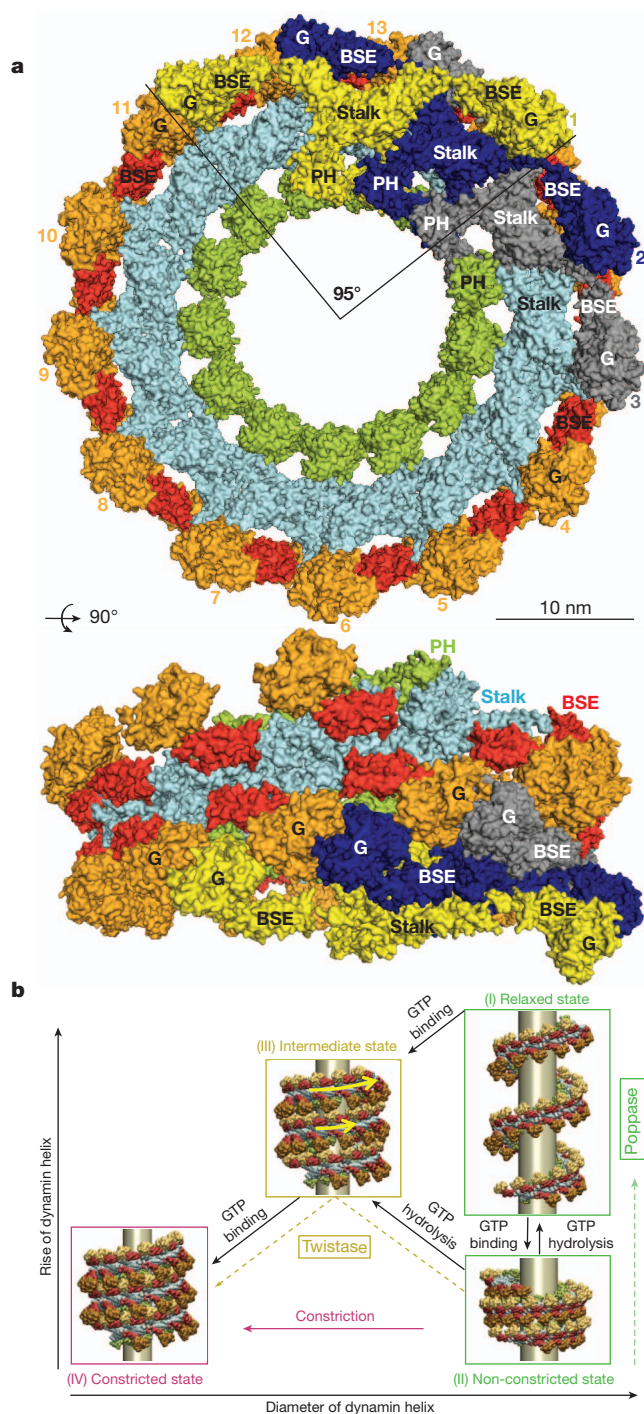


Figure 4 | Model for dynamin oligomerization and function. **a**, Model of the oligomerized dynamin helix in the constricted state, in top and side view (see also Supplementary Fig. 10). Three dimers (1–3) are uniformly coloured. Whereas 13 stalk dimers complete one turn, the G domain of dimer (*i*) associates with the G domain of dimer (*i* + 10). **b**, Structure-based illustration for the proposed mechanism of the dynamin oligomer. Variations in the assembly of consecutive dynamin molecules lead to dynamin helices of different rise and diameter. For further explanation, see main text and Supplementary Fig. 12.

around hinge 1 upon oligomerization. Additionally, the G domain separates from the BSE by a rotation around the invariant Pro 32 and Pro 294 (Supplementary Fig. 6). The corresponding residues Gly 68 and Gly 309 in BDL^{33,35} and Pro 342 in atlastin^{36,37} have also been suggested to act as a hinge (hinge 2). Integrity of the G domain–BSE interface is crucial for the function of dynamin, as indicated by the

aggregation of the I10D interface mutant *in vivo* (Supplementary Figs 6 and 8). Furthermore, I10D behaved as a dominant-negative mutant in transferrin uptake assays (Fig. 3b). Moreover, the QD17–18AA mutation rendered dynamin 2 largely inactive in endocytosis assays, despite localizing correctly to clathrin-coated pits at the plasma membrane (Fig. 3b and Supplementary Fig. 8).

The PH domains also undergo a pronounced rearrangement, to a position below the stalk, with the lipid-binding loops^{23–25} oriented towards the membrane (Supplementary Fig. 11a). We suggest that the stalk PH domain interface is disrupted by binding of the PH domain to high-affinity, phosphatidylinositol-4,5-bisphosphate-containing membranes such as the plasma membrane^{23–25}, thereby promoting dynamin oligomerization.

Finally, we suggest that rotation of stalk dimers via the flexible interface-1 and/or interface-3, leaving interface-2 unchanged, leads to bending of the linear oligomer and allows helix formation (Supplementary Figs 11b and 12). Interestingly, the IHGIR395–399AAAAA mutation in interface-3 prevented liposome binding and, consequently, the liposome-stimulated GTPase activity (Supplementary Fig. 7). This mutant further behaved in a dominant-negative fashion in transferrin uptake assays (Fig. 3b) and displayed a diffuse localization and reduced recruitment to clathrin-coated pits (Supplementary Fig. 8). These results point to the central role of interface-3 for the function of dynamin.

Discussion

The present work, combined with prior studies^{18–20,38}, suggests a structural model for the mechanochemical coupling in dynamin that is consistent with previous models^{9–11}. Our structural analysis indicates that the diameter of helical dynamin assemblies is controlled by the angle between two stalk dimers (Supplementary Figs 11b and 12f–h). We suggest that this angle is adjusted in response to GTP-dependent dimerization of the G domains: a relaxed conformation of interface-1 is adopted in the absence of G domain constraints, whereas GTP-triggered dimerization of the G domains constrains rotation in interface-1 and induces a bent, constricted conformation (Supplementary Fig. 11b), possibly via the BSE–stalk interface (Fig. 3a). Interestingly, in molecular dynamics simulations of two stalk dimers, the bent conformation of interface-1 rapidly converts towards a relaxed state, concomitant with an opening of the helical oligomer (Fig. 4b and Supplementary Fig. 12). This supports our assumption that constraints from G domains dimerization are required for the stabilization of the constricted state.

Accordingly, dynamin initially assembles via the stalks with interface-1 in a relaxed conformation, allowing the filament to adopt a range of different diameters¹² (state I in Fig. 4b and Supplementary Fig. 12). When the filament has embraced its template, GTP-loaded G domains of adjacent turns dimerize, the inhibitory stalk–BSE interface (Fig. 3a) is disrupted and the bent conformation of interface-1 is induced. This will result in constriction of the filament if the lipid template is flexible (state IV in Fig. 4b, constrictase model¹⁰) or compaction of the dynamin helix if the lipid template is rigid (state II in Fig. 4b, poppase model¹¹). Constriction of a long dynamin helix will induce a sliding of neighbouring filaments until a new constricted equilibrium position of the oligomer is reached. This sliding is observed as a rotary movement of the dynamin helix upon addition of GTP in real-time imaging assays (twistase model⁹). To reach the constricted state along the whole assembly, several cycles of local release and rebinding of neighbouring dynamin turns might be triggered by GTP-dependent dimerization of G domains and dissociation after GTP hydrolysis (state III in Fig. 4b). Accordingly, GTP binding and hydrolysis are both required for the mechanochemical function of dynamin^{8,39} and might induce local opening or twisting of the constricted dynamin helix. The resulting shear forces could tear the underlying membrane.

The complex domain interplay in dynamin rationalizes the high degree of structural conservation in the dynamin superfamily, sheds light on the molecular mechanisms of disease-associated mutations and highlights structural features of the nucleotide-free state as a prerequisite to understand dynamin's mechanochemical function.

METHODS SUMMARY

A human dynamin 1 construct (residues 6–746) containing the IHGIR395–399AAAAA and K562E mutations was expressed as a His₆-tag-fusion in *Escherichia coli* and purified to homogeneity. Crystals were obtained using PEG400 and isopropanol as precipitants. The structure was solved by molecular replacement. To verify the model, the positions of 19 out of 26 methionines were determined by an anomalous data set of SeMet-substituted crystals. Liposome binding assays were carried out as described (<http://www.endocytosis.org>). GTPase assays were carried out at 37 °C using 20 mM HEPES/NaOH, pH 7.5, 150 mM NaCl, 2 mM KCl, 2 mM MgCl₂ as reaction buffer. Uptake of fluorescently labelled transferrin was monitored in HeLa cells depleted of endogenous dynamin 2 by siRNA and transfected with the indicated siRNA-resistant dynamin 2 constructs.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 8 March; accepted 17 July 2011.

Published online 18 September 2011.

1. Praefcke, G. J. & McMahon, H. T. The dynamin superfamily: universal membrane tubulation and fission molecules? *Nature Rev. Mol. Cell Biol.* **5**, 133–147 (2004).
2. van der Bliek, A. M. & Meyerowitz, E. M. Dynamin-like protein encoded by the *Drosophila shibire* gene associated with vesicular traffic. *Nature* **351**, 411–414 (1991).
3. Ferguson, S. M. *et al.* A selective activity-dependent requirement for dynamin 1 in synaptic vesicle endocytosis. *Science* **316**, 570–574 (2007).
4. Robinson, P. J. *et al.* Dynamin GTPase regulated by protein kinase C phosphorylation in nerve terminals. *Nature* **365**, 163–166 (1993).
5. Lu, J. *et al.* Postsynaptic positioning of endocytic zones and AMPA receptor cycling by physical coupling of dynamin-3 to Homer. *Neuron* **55**, 874–889 (2007).
6. Cook, T. A., Urrutia, R. & McNiven, M. A. Identification of dynamin 2, an isoform ubiquitously expressed in rat tissues. *Proc. Natl Acad. Sci. USA* **91**, 644–648 (1994).
7. Durieux, A. C. *et al.* Dynamin 2 and human diseases. *J. Mol. Med.* **88**, 339–350 (2010).
8. Marks, B. *et al.* GTPase activity of dynamin and resulting conformation change are essential for endocytosis. *Nature* **410**, 231–235 (2001).
9. Roux, A. *et al.* GTP-dependent twisting of dynamin implicates constriction and tension in membrane fission. *Nature* **441**, 528–531 (2006).
10. Hinshaw, J. E. & Schmid, S. L. Dynamin self-assembles into rings suggesting a mechanism for coated vesicle budding. *Nature* **374**, 190–192 (1995).
11. Stowell, M. H. B. *et al.* Nucleotide-dependent conformational changes in dynamin: evidence for a mechanochemical molecular spring. *Nature Cell Biol.* **1**, 27–32 (1999).
12. Bashkurov, P. V. *et al.* GTPase cycle of dynamin is coupled to membrane squeeze and release, leading to spontaneous fission. *Cell* **135**, 1276–1286 (2008).
13. Pucadyil, T. J. & Schmid, S. L. Real-time visualization of dynamin-catalyzed membrane fission and vesicle release. *Cell* **135**, 1263–1275 (2008).
14. Timm, D. *et al.* Crystal structure of the pleckstrin homology domain from dynamin. *Nature Struct. Biol.* **1**, 782–788 (1994).
15. Ferguson, K. M. *et al.* Crystal structure at 2.2 Å resolution of the pleckstrin homology domain from human dynamin. *Cell* **79**, 199–209 (1994).
16. Niemann, H. H. *et al.* Crystal structure of a dynamin GTPase domain in both nucleotide-free and GDP-bound forms. *EMBO J.* **20**, 5813–5821 (2001).
17. Reubold, T. F. *et al.* Crystal structure of the GTPase domain of rat dynamin 1. *Proc. Natl Acad. Sci. USA* **102**, 13093–13098 (2005).
18. Mears, J. A., Ray, P. & Hinshaw, J. E. A corkscrew model for dynamin constriction. *Structure* **15**, 1190–1202 (2007).
19. Chappie, J. S. *et al.* G domain dimerization controls dynamin's assembly-stimulated GTPase activity. *Nature* **465**, 435–440 (2010).

20. Gao, S. *et al.* Structural basis of oligomerization in the stalk region of dynamin-like MxA. *Nature* **465**, 502–506 (2010).
21. Ramachandran, R. *et al.* The dynamin middle domain is critical for tetramerization and higher-order self-assembly. *EMBO J.* **26**, 559–566 (2007).
22. Chappie, J. S. *et al.* An intramolecular signaling element that modulates dynamin function *in vitro* and *in vivo*. *Mol. Biol. Cell* **20**, 3561–3571 (2009).
23. Vallis, Y. *et al.* Importance of the pleckstrin homology domain of dynamin in clathrin-mediated endocytosis. *Curr. Biol.* **9**, 257–263 (1999).
24. Zheng, J. *et al.* Identification of the binding site for acidic phospholipids on the pH domain of dynamin: implications for stimulation of GTPase activity. *J. Mol. Biol.* **255**, 14–21 (1996).
25. Salim, K. *et al.* Distinct specificity in the recognition of phosphoinositides by the pleckstrin homology domains of dynamin and Bruton's tyrosine kinase. *EMBO J.* **15**, 6241–6250 (1996).
26. Zhang, P. & Hinshaw, J. E. Three-dimensional reconstruction of dynamin in the constricted state. *Nature Cell Biol.* **3**, 922–926 (2001).
27. Kenniston, J. A. & Lemmon, M. A. Dynamin GTPase regulation is altered by PH domain mutations found in centronuclear myopathy patients. *EMBO J.* **29**, 3054–3067 (2010).
28. Kochs, G. *et al.* Assay and functional analysis of dynamin-like Mx proteins. *Methods Enzymol.* **404**, 632–643 (2005).
29. Ingerman, E. *et al.* Dnm1 forms spirals that are structurally tailored to fit mitochondria. *J. Cell Biol.* **170**, 1021–1027 (2005).
30. Chang, C. R. *et al.* A lethal *de novo* mutation in the middle domain of the dynamin-related GTPase Drp1 impairs higher order assembly and mitochondrial division. *J. Biol. Chem.* **285**, 32494–32503 (2010).
31. Graham, M. E. *et al.* The *in vivo* phosphorylation sites of rat brain dynamin I. *J. Biol. Chem.* **282**, 14695–14707 (2007).
32. Rush, J. *et al.* Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nature Biotechnol.* **23**, 94–101 (2004).
33. Low, H. H. & Lowe, J. Dynamin architecture—from monomer to polymer. *Curr. Opin. Struct. Biol.* **20**, 791–798 (2010).
34. Wang, L. *et al.* Dynamin 2 mutants linked to centronuclear myopathies form abnormally stable polymers. *J. Biol. Chem.* **285**, 22753–22757 (2010).
35. Low, H. H. *et al.* Structure of a bacterial dynamin-like protein lipid tube provides a mechanism for assembly and membrane curving. *Cell* **139**, 1342–1352 (2009).
36. Byrnes, L. J. & Sonderrmann, H. Structural basis for the nucleotide-dependent dimerization of the large G protein atlastin-1/SPG3A. *Proc. Natl Acad. Sci. USA* **107**, 10173/pnas.1012792108 (10 January 2011).
37. Bian, X. *et al.* Structures of the atlastin GTPase provide insight into homotypic fusion of endoplasmic reticulum membranes. *Proc. Natl Acad. Sci. USA* **108**, 3976–3981 (2011).
38. Klockow, B. *et al.* The dynamin A ring complex: molecular organization and nucleotide-dependent conformational changes. *EMBO J.* **21**, 240–250 (2002).
39. Morlot, S. *et al.* Deformation of dynamin helices damped by membrane friction. *Biophys. J.* **99**, 3580–3588 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This project was supported by a grant from the Deutsche Forschungsgemeinschaft (SFB 740/C7 and SFB958/A12 to O.D., SFB740/D7 and SFB958/A04 to F.N., SFB740/C8 and SFB 958/A7 to V.H.), by a Career Development Fellowship of The International Human Frontier Science Program Organization and by the EMBO Young Investigator Program to O.D. We would like to thank S. Werner, M. Papst and S. Kraft for technical assistance, the BESSY staff at BL14.1 for help during data collection, G. Schröder for advice in DEN refinement and U. Heinemann for discussions.

Author Contributions K.F., Y.P., D.S. and Y.R. performed experiments, K.F., Y.P., S.G., M.H., V.H., F.N. and O.D. designed research, M.H. and F.N. conducted and analysed molecular modelling and molecular dynamics simulations. K.F. and O.D. wrote the manuscript.

Author Information The atomic coordinates of human dynamin 1 have been deposited in the Protein Data Bank with accession number 3SNH. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to O.D. (oliver.daumke@mdc-berlin.de) or K.F. (katja.faelber@mdc-berlin.de).

METHODS

Protein expression and purification. Human dynamin 1 (residues 6–746) and indicated mutants of this construct were expressed from pET46-EK/LIC (Novagen) as N-terminal His₆-tag fusion followed by a PreScission cleavage site. The crystallized construct contained mutations IHGIR395–399AAAAA which prevented oligomerization and K562E which reduced DNA contaminations during purification. Proteins were expressed in *Escherichia coli* host strain Rosetta2-BL21-DE3, and bacteria were cultured in autoinduction medium (Novagen) for 7.5 h at 37 °C followed by temperature shift to 20 °C for overnight expression. Selenomethionine (SeMet)-substituted human dynamin 1 was expressed in M9 minimal medium, supplemented with L-amino acids Lys, Phe, Thr (100 mg l⁻¹), Ile, Leu, Val, SeMet (50 mg l⁻¹), using the same vector and host strain as for native protein expression⁴⁰. Cells were resuspended in buffer A (25 mM HEPES/NaOH (pH 7.9), 500 mM NaCl, 2 mM MgCl₂, 1 μM DNase (Roche), 500 μM Pefabloc (Roth)) and disrupted by a microfluidizer (Microfluidics). Cleared lysates (95,000g, 1 h, 4 °C) were incubated with Benzonase (Novagen) for 30 min at 4 °C before application to a Co²⁺-Talon column (Clontech). Protein was eluted with buffer A containing additional 100 mM imidazole. Fractions containing human dynamin 1 were incubated with 2.4 mM β-mercaptoethanol and His₆-tagged PreScission protease overnight at 4 °C. Using 50 kDa molecular weight cut-off concentrators (Amicon), imidazole, β-mercaptoethanol and the free His₆-tag were removed by washing with buffer A, before a second application to a Co²⁺-Talon column to remove non-cleaved His-tagged dynamin 1 and protease. The flow-through and four column volumes of washing buffer A were collected and concentrated. Finally, dynamin 1 was purified by size exclusion chromatography on a Superdex200 column (GE) in buffer containing 25 mM HEPES/NaOH (pH 7.5), 300 mM NaCl, 2 mM MgCl₂. Fractions containing dynamin 1 were pooled, concentrated and flash-frozen in liquid nitrogen (Supplementary Fig. 1). SeMet-substituted protein was purified using the same protocol.

Crystallization and structure determination. Crystallization trials by the sitting-drop vapour-diffusion method were performed at 4 °C using a Hydra-plus-One pipetting robot (Matrix Technologies Corporation) and Rock Imager storage system (Formulatrix). The human dynamin 1 construct (300 nl at a concentration of 12 mg ml⁻¹) was mixed with an equal volume of reservoir solution containing 9% PEG400, 6% isopropanol, 100 mM HEPES/NaOH (pH 7.3), 10 mM MgCl₂ and 10 mM KCl. Crystals of the native protein appeared after 2 weeks and had dimensions of 0.2 mm × 0.05 mm × 0.02 mm. Crystals of SeMet protein were obtained in 6% MPD, 10% isopropanol, 0.1 M HEPES/NaOH (pH 7.5). During flash-cooling of the crystals in liquid nitrogen, a cryo-solution containing additionally 20% ethylene glycol was used. All data were recorded at BL14.1 at BESSY II, Berlin. One native data set was collected from a single crystal and processed and scaled using the XDS program suite⁴¹. Phase information was obtained by molecular replacement with Phaser⁴² using the structure of the nucleotide-free rat dynamin 1 G domain¹⁶, the stalk of human Mx²⁰ and the human PH domain¹⁵ as search models. The model was built using COOT⁴³ and iteratively refined using CNS1.3 with a deformable elastic network⁴⁴, followed by 10 cycles of TLS refinement using 1 TLS group per domain in refmac5 (ref. 45). To confirm the amino acid sequence, a data set of a SeMet-substituted protein crystal was recorded at the selenium peak wavelength. Owing to non-isomorphism of the native and SeMet crystals, molecular replacement with the refined native structure was performed. The positions of 19 out of 26 selenium atoms in the dynamin 1 construct were determined by calculating an anomalous difference map with the CCP4 program suite⁴⁶ using the calculated phases after refinement. The final native model comprises amino acids 6–746. Residues 63–64, 110–112, 143–149, 347–356, 394–404, 446–447, 500–517, 534–537, 578–581 and 632–652 are disordered. Electron density for loop L2^{BS} (amino acids 709–715) was only visible in the SeMet derivative structure and was therefore not modelled in the native structure. However, L2^{BS} was included for figure preparation. Of all residues, 86.9% are in the most favoured regions of the Ramachandran plot and no residue in the disallowed regions, as analysed with Procheck⁴⁷. Figures were prepared with PyMol⁴⁸. Buried surface areas (per molecule) were calculated using CNS⁴⁴. Solvent-accessible areas per residue were calculated using areaimol⁴⁶. Domain superpositions were performed with lsqkab⁴⁶. Sequences were aligned using CLUSTAL W⁴⁹ and adjusted by hand.

Liposome co-sedimentation assays. Liposomes were prepared as previously described (<http://www.endocytosis.org>). Folch liposomes (total bovine brain lipids fraction I from Sigma) in 20 mM HEPES/NaOH (pH 7.5), 150 mM NaCl, 2 mM KCl were extruded through a 0.1 μm filter. Liposomes (0.1 mg ml⁻¹) were incubated at room temperature with 4.0 μM of the indicated dynamin 1 construct for 10 min in 40 μl reaction volume, followed by a 70,000g spin for 10 min at 20 °C.

GTP hydrolysis assay. GTPase activities of 1 μM of the indicated dynamin 1 constructs were determined at 37 °C in 20 mM HEPES/NaOH (pH 7.5), 150 mM

NaCl, 2 mM KCl, 2 mM MgCl₂, in the absence and presence of 0.1 mg ml⁻¹ 0.1-μm-filtered Folch liposomes, using saturating concentrations of GTP as substrate (1 mM for the basal and 3 mM for the stimulated reactions). Reactions were initiated by the addition of protein to the reaction. At different time points, reaction aliquots were diluted 15-fold in GTPase buffer and quickly transferred to liquid nitrogen. Nucleotides in the samples were separated via a reversed-phase Hypersil ODS-2 C18 column (250 × 4 mm), with 10 mM tetrabutylammonium bromide, 100 mM potassium phosphate (pH 6.5), 7.5% acetonitrile as running buffer. Denatured proteins were adsorbed at a C18 guard column. Nucleotides were detected by absorption at 254 nm and quantified by integration of the corresponding peaks. Rates were derived from a linear fit to the initial reaction (<40% GTP hydrolysed).

Transferrin uptake in HeLa cells. HeLa cells were transfected with siRNA using Oligofectamine (Invitrogen) on day 1. The sequence of the siRNA against human dynamin 2 is 5'-GCAACUGACCAACCACAUC-3', targeting nucleotides 849–867. On day 2, cells were transfected with pEGFP-N1 (Clontech) or siRNA-resistant rat dynamin 2-pEGFP-N1 using lipofectamine 2000 (Invitrogen) and seeded on coverslips. On day 3, cells were serum-starved and incubated with 20 μg ml⁻¹ transferrin-Alexa568 (Molecular Probes, Invitrogen) for 10 min at 37 °C. After four PBS washes on ice, cells were paraformaldehyde-fixed for 20 min at room temperature. Transferrin uptake was analysed using a Zeiss Axiovert200M microscope and Slidebook imaging software (3i Inc.). Internalized transferrin was quantified from transfected cells only and normalized to the value of eGFP-transfected, mock-treated cells (*n* = 28–83 images, five independent experiments; IHGVR395–399AAA, 110D, QD17–18AA: three independent experiments; E368K, R369W, S619L, F372D: two independent experiments). Knockdown of dynamin 2 and expression levels of dynamin 2-eGFP mutants were assessed by immunoblotting using antibodies to dynamin 2 (a gift of M. A. McNiven), β-actin (clone ac-15, Sigma-Aldrich) and eGFP (clone j1-8, Clontech).

Localization of dynamin 2-eGFP mutants. HeLa cells were transfected with dynamin 2-pEGFP-N1 wild-type or mutant constructs 20 h before fixation in 4% paraformaldehyde for 12 min at room temperature. Cells were blocked and permeabilized in 10% goat serum, 0.3% Triton X-100, 100 mM NaCl in phosphate buffer and stained for endogenous adaptor complex AP-2 using an antibody to α-adaptin (clone AP-6, Abcam). Total internal reflection fluorescence (TIRF) imaging was performed using a Zeiss Axiovert200M microscope equipped with a ×100 TIRF objective and a dual-colour TIRF setup from Visitron Systems using Slidebook imaging software (3i Inc.).

Loop modelling and molecular dynamics simulations. For modelling of the unresolved loop regions L1N^S and L2^S, two stalk dimers in the constricted state served as scaffold. Using Modeller (9v8)⁵⁰, the scaffold was fixed in position, whereas L1N^S and L2^S could freely sample the empirical potential function. To reduce the conformational search space, additional harmonic distance restraints were added between conserved residues Arg 399–Asp 406 and Glu 355–Arg 361. Based on the modelled stalk tetramer, five independent all-atom molecular dynamics simulations (NVT ensemble), each of 90 ns, were conducted at *T* = 300 K in a periodic boundary setting using Gromacs 4.5.3 (ref. 51). The model was immersed in a rectangular 20 nm × 10 nm × 9 nm box, containing approximately 56,400 water molecules, 21 sodium and 17 chloride ions to neutralize the system, resulting in a total number of 185,857 atoms. As force fields, Amber99 (protein and ions)⁵² and TIP3P (water)⁵³ were applied. To treat long-range interactions, the Particle-mesh Ewald method^{54,55} was used. A cut-off of 1 nm was used for the real parts of electrostatic and van der Waals interactions. All hydrogen bonds were constrained by using the LINCS⁵⁶ algorithm, allowing for an integration time-step of 2 fs. For the thermostatted integration, Langevin dynamics were used as implemented by the Gromacs s.d. integrator (tau_t = 1).

For the calculation of bending and twisting angles, each of the four stalk monomers was represented by two geometric centres, defined as the mean position of Cα atoms of residues 366–377, 420–430, 468–481 and 671–683 for position A and residues 360–365, 428–445, 457–472 and 686–701 for position B (Supplementary Fig. 12d). The stalk bending angle α was defined as the mean angle between parallel stalks, and the twisting angle β by the minimal angle between the planes spanned by each dimer (positions A, B, B' in Supplementary Fig. 12f).

For each simulation time step, the corresponding stalk tetramer structure describes a linear transformation of the first dimer onto the second dimer, consisting of a translation vector and a rotation matrix. This linear transformation was used to reconstruct the structure of an ideal dynamin helix by applying it to the dynamin dimer model in the constricted state. The diameter, *d*, and the rise per turn, *r*, of these helices were measured by using the geometric centres of the stalk coordinates and obtaining trajectories in (*d*, *r*). Based on these trajectories, the free energy surface of stalk helix conformations was calculated: the two-dimensional space (*d*, *r*) was discretized into boxes of size 25 × 25 Å. Based on the simulation trajectories, the transition probability between all pairs of boxes was computed,

which allowed the calculation of an equilibrium probability of finding a single stalk tetramer in a given box, $p_1(d,r)^{57}$. When more than two stalk dimers are assembled, non-cooperative behaviour of neighbouring dimers has to be considered, for example, the assembled stalk dimers can almost independently switch between different conformations. The resulting equilibrium distribution of two independent tetrameric units would be given by the convolution of two single-tetramer distributions, $p_2(d,r)$. It was found that for only about three such convolutions, the resulting probability distribution converges to $p_3(d,r) \approx p(d,r)$. Thus, assuming that the helix has at least three independently switching subunits, the free energy landscape is unique, and is given by $F(d,r) = -kT \ln(p(d,r))$, where k is the Boltzmann constant and T the temperature.

40. Van Duyne, G. D. *et al.* Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J. Mol. Biol.* **229**, 105–124 (1993).
41. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.* **26**, 795–800 (1993).
42. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
43. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
44. Schröder, G. F., Levitt, M. & Brunger, A. T. Super-resolution biomolecular crystallography with low-resolution data. *Nature* **464**, 1218–1222 (2010).
45. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
46. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
47. Laskowski, R. A. *et al.* PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291 (1993).
48. DeLano, W. L. *The PyMol Molecular Graphics System* version 1.4.1 (Schrödinger, 2011).
49. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
50. Fiser, A., Do, R. K. & Sali, A. Modeling of loops in protein structures. *Protein Sci.* **9**, 1753–1773 (2000).
51. Van Der Spoel, D. *et al.* GROMACS: fast, flexible, and free. *J. Comput. Chem.* **26**, 1701–1718 (2005).
52. Wang, J. M., Cieplak, P. & Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **21**, 1049–1074 (2000).
53. Jorgensen, W. *et al.* Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
54. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995).
55. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: an $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).
56. Hess, B. *et al.* LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).
57. Prinz, J. H. *et al.* Markov models of molecular kinetics: generation and validation. *J. Chem. Phys.* **134**, 174105 (2011).

The crystal structure of dynamin

Marijn G. J. Ford¹, Simon Jenni² & Jodi Nunnari¹

Dynamin-related proteins (DRPs) are multi-domain GTPases that function via oligomerization and GTP-dependent conformational changes to play central roles in regulating membrane structure across phylogenetic kingdoms. How DRPs harness self-assembly and GTP-dependent conformational changes to remodel membranes is not understood. Here we present the crystal structure of an assembly-deficient mammalian endocytic DRP, dynamin 1, lacking the proline-rich domain, in its nucleotide-free state. The dynamin 1 monomer is an extended structure with the GTPase domain and bundle signalling element positioned on top of a long helical stalk with the pleckstrin homology domain flexibly attached on its opposing end. Dynamin 1 dimer and higher order dimer multimers form via interfaces located in the stalk. Analysis of these interfaces provides insight into DRP family member specificity and regulation and provides a framework for understanding the biogenesis of higher order DRP structures and the mechanism of DRP-mediated membrane scission events.

Dynamin-related proteins (DRPs) belong to a highly conserved (Supplementary Fig. 1) GTPase superfamily that catalyses diverse membrane remodelling events^{1,2}. Membrane scission DRPs include dynamin 1, which catalyses clathrin-coated vesicle scission at the plasma membrane and Drp1/Dnm1, which divide mitochondria. Despite their functional diversity, all DRPs undergo GTP cycle-dependent conformational changes to regulate self-assembly and disassembly^{3–8}. DRP architecture includes an amino-terminal GTPase domain, a bundle signalling element (BSE), a middle domain (MD) and a GTPase effector domain (GED)^{9,10}. Many DRPs also have a variable region between the MD and GED; in dynamin 1, this is a pleckstrin homology domain (PH) that binds to phosphatidylinositol-4,5-bisphosphate (PtdIns-4,5-P₂)-containing membranes to facilitate targeting and possibly membrane remodelling via membrane insertion^{11–14}.

Assembled DRPs can form helical structures *in vitro*^{15,16}. Within these structures, GTP cycle-driven conformational changes result in membrane remodelling^{4,5,7,17}. The structural basis for DRP self-assembly and GTP cycle-dependent conformational changes are not fully understood. The mechanism of assembly has been informed by the structure of the 'stalk' of MxA, a distantly related DRP, in which the MD and part of the GED form an extended helical bundle that mediates self-assembly via conserved interfaces¹⁸. Several structures have also provided insight into DRP GTP cycle conformational changes (Supplementary Fig. 2)^{19–24}. Cryo-electron microscopic structures of assembled dynamin in guanosine-5'-[(β,γ)-methylene]triphosphate (GMPPCP)-bound and nucleotide-free states have provided models for the assembled oligomers and the location of the GTPase and PH domains within the helical structures^{17,25}. However, to understand the basis of DRP self-assembly and mechanism, the architecture of DRP domains within a single molecule must be elucidated. Here, we report the crystal structure of an assembly-deficient dynamin 1 in the nucleotide-free state that lacks only the unstructured carboxy-terminal proline-rich domain (PRD).

G397D, an assembly-deficient dynamin 1 mutant

The propensity of DRPs to assemble has been an obstacle to obtaining crystals suitable for diffraction experiments. We identified the mutation G436D, an invariant MD residue, in the *Saccharomyces cerevisiae* mitochondrial division DRP Dnm1, by screening for mutations that possessed the same phenotype as assembly-deficient Dnm1 G385D,

specifically mutations that shifted Dnm1–green fluorescent protein (GFP) fusion from a punctate to a more diffuse localization pattern in yeast cells¹⁶. We expressed and purified (Supplementary Fig. 3a) the orthologous rat dynamin 1 mutant protein lacking the PRD (Dyn1 G397D Δ PRD) and examined its ability to self assemble using light scattering (Fig. 1a). Addition of GMPPCP caused an increase in scattering in Dyn1 Δ PRD samples, but caused no change in scattering in Dyn1 G397D Δ PRD samples. Using a combination of size-exclusion chromatography and sucrose gradient centrifugation²⁶, Dyn1 Δ PRD and Dyn1 G397D Δ PRD were estimated to be dimeric (Supplementary Fig. 3b and Supplementary Table 1) under non-assembly conditions, similar to other assembly-deficient DRP mutants^{16,27}. Under assembly conditions, we observed an increase in the sedimentation coefficient of Dyn1 Δ PRD, but not for Dyn1 G397D Δ PRD (Supplementary Table 1). Dyn1 G397D Δ PRD was defective for assembly-stimulated GTP hydrolysis (Fig. 1b and Supplementary Table 2) and failed to assemble into helical structures on phosphatidylinositol 4-phosphate (PtdIns-4P)-containing lipid nanotubes, in contrast to Dyn1 Δ PRD (Fig. 1c). Dyn1 G397D Δ PRD was also soluble at higher concentrations than Dyn1 Δ PRD. Together, these data indicate that the G397D mutation severely hampers dynamin self assembly and substantiates the critical role of the MD in intermolecular interactions^{16,18,27–29}. Given these characteristics, it presented an attractive target for crystallization.

Crystallization and structure of Dyn1 G397D Δ PRD

We obtained orthorhombic crystals of Dyn1 G397D Δ PRD that diffracted to a minimum Bragg spacing of 3.1 Å (Methods and Supplementary Table 3). The structure was solved by molecular replacement, using the nucleotide-free rat dynamin 1 GTPase domain, the human dynamin 1 PH and a portion of the human MxA stalk, as sequential search models (Methods)^{11,13,18,30}. We traced the complete model, except for some disordered loops, and assigned amino acids after refinement of the molecular replacement solution. The model was refined to R/R_{free} values of 0.21/0.27. A representative example of the B-sharpened likelihood-weighted $2mF_{\text{obs}} - DF_{\text{calc}}$ electron density map is shown in Fig. 2c.

Dyn1 G397D Δ PRD forms an extended structure with the GTPase and PH domains separated by a stalk consisting of an anti-parallel helical bundle of the MD and a helix from the GED (Fig. 2a, b). By

¹Department of Molecular and Cellular Biology, University of California, Davis, Davis, California 95616, USA. ²Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA.

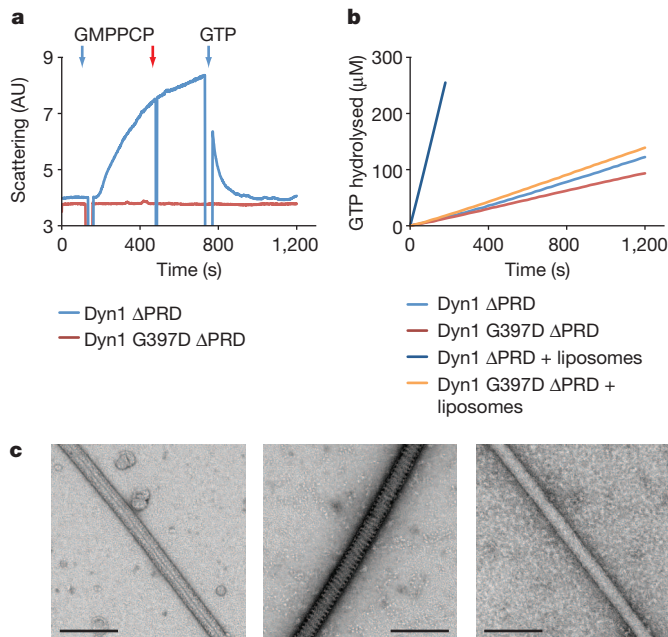


Figure 1 | The G397D mutation in Dyn1 Δ PRD blocks self-assembly. **a**, 90° light scattering. Dyn1 Δ PRD (blue, 1 μ M) and Dyn1 G397D Δ PRD (red, 1 μ M) were monitored after addition of 0.5 mM GMPPCP and, in the case of Dyn1 Δ PRD, 1 mM GTP (blue arrows). Red arrow indicates opening of sample port. AU, arbitrary units. **b**, Steady-state GTP hydrolysis kinetics of Dyn1 Δ PRD and Dyn1 G397D Δ PRD in the absence (light blue and red) and presence (dark blue and orange) of 0.1 mg ml⁻¹ liposomes containing 10% PtdIns-4P, monitored by a NADH-dependent coupled assay as described⁴³. A representative trace is shown with 1 mM GTP. **c**, Transmission electron microscopy of negative-stained 0.25 mg ml⁻¹ 10% PtdIns-4P lipid nanotubes with Dyn1 Δ PRD (middle) and Dyn1 G397D Δ PRD (right) and no added protein (left) and 0.5 mM GMPPCP. Scale bars, 200 nm.

contrast, the distantly related bacterial dynamin-like protein (BDLP) is compactly folded in its nucleotide-free and GDP-bound states (Supplementary Fig. 2a), with its 'paddle', in an analogous region to the PH domain of dynamin, in close proximity to the GTPase domain^{19,20}. Thus, unlike BDLP, GTP binding in dynamin is not harnessed to form an extended structure, consistent with nucleotide-independent assembly of dynamin on liposomes^{4,8,31,32}. Linkers connecting the MD to the PH and the PH to the GED are disordered, indicating a flexibly tethered PH. Thus, any of three crystallographic symmetry-related PH domains could connect to the remainder of the structure (Fig. 2d). The probable PH partner, based on the fit with the envelope of an assembly-deficient dynamin dimer determined by small-angle X-ray scattering³³ (SAXS), is shown in Fig. 2b. The structure of the PH is similar to those previously determined^{11,13}, with expected differences concentrated in the variable loops.

The structure of the GTPase domain is similar to that of the previously determined nucleotide-free dynamin GTPase domain structure³⁰. There are minor expected changes in the poorly resolved switch 2 region, the dynamin-specific loop and in the loop connecting the N_{GTPase} to the GTPase domain. The C_{GTPase} helix is kinked at the conserved proline 294 and, together with the N_{GTPase} helix and a helix at the C terminus of the GED (C_{GED}), forms the three-helix BSE⁹. C_{GED} covers a groove of 937 Å² between the N_{GTPase} and C_{GTPase} helices. In the nucleotide-free dynamin 1 GTPase and the GDP-bound *Dictyostelium discoideum* dynamin A GTPase structures, a myosin peptide substitutes for C_{GED}, indicating the importance of this interface for GTPase domain stability^{30,34}.

A linker with elevated crystallographic temperature factors connects the C_{GTPase} to the dynamin stalk (Fig. 2a, b) and contains two prolines (319 and 322), whose equivalents in the distantly related DRP atlastin 1 connect the GTPase and MD (Supplementary Fig. 2b).

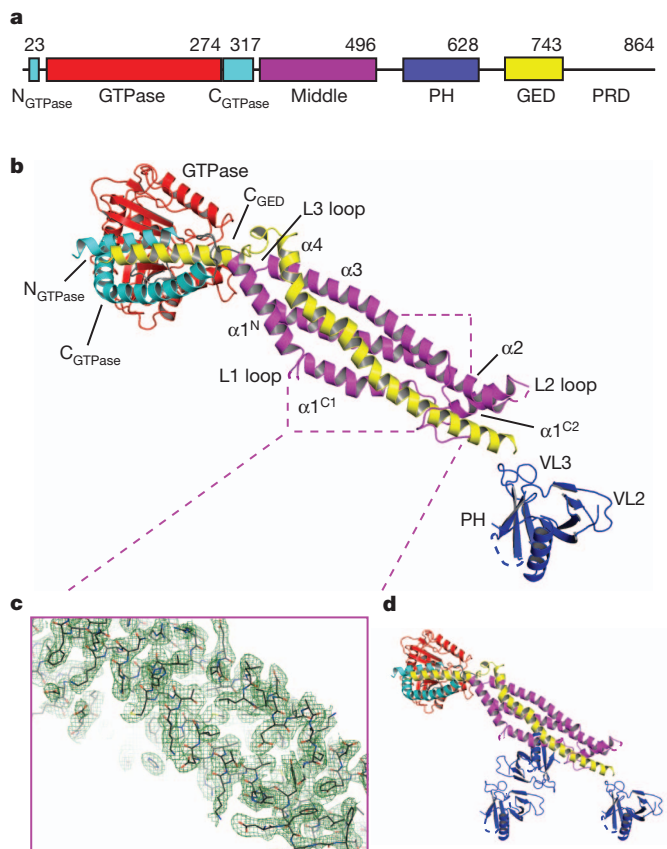


Figure 2 | The crystal structure of Dyn1 G397D Δ PRD. **a**, Schematic of the Dyn1 domain structure. Numbers indicate domain-ending amino acid. Colour scheme used here is retained. **b**, Crystal structure of Dyn1 G397D Δ PRD. Linkers between the N_{GTPase} and GTPase domain and the C_{GTPase} and MD are shown in grey. Loops with no density are represented with dashed lines. Stalk nomenclature is based on that of the MxA stalk structure¹⁸. VL, variable loop. **c**, An example of the refined to B-sharpened $2mF_{\text{obs}} - DF_{\text{calc}}$ map, contoured at 1 σ . The region shown is part of the stalk boxed by purple dotted lines in **(b)**. **d**, Three symmetry-related PH domains in the lattice.

Although the dynamin stalk sequence shares limited identity to MxA, its overall structure is similar and the MxA nomenclature is retained¹⁸ (Supplementary Fig. 4a). As for MxA, helix α 1 is split by a disordered loop, L1. The remainder of α 1 diverges from MxA and is split into two helices, termed α 1^{C1} and α 1^{C2}. Helix α 1^{C2} connects to helix α 2 via a short disordered loop, L2. Helices α 2 and α 3 run the length of the stalk and are joined by a short loop, L3. The stalk is completed by GED-derived helix α 4 that spans the stalk and connects to C_{GED} via a linker. Following helix α 3 is a coil that folds across α 4 that is strongly conserved in dynamins, Drp1s and Dnm1, but absent in MxA.

The dynamin–dimer interface

The crystal lattice contains linear filaments of dynamin assembled via three stalk interfaces, similar to MxA¹⁸, resulting in layers of interacting stalks separated by GTPase and PH domains (Fig. 3a, b). Interface 2, the largest with a buried surface of 1,339 Å², has two-fold symmetry and is formed by residues from stalk helices α 4 and α 3 (Fig. 3c), with an additional residue from α 1^{C1} (H367). Each protomer in the interface contributes seven direct hydrogen bonds and eight hydrophobic residues line the site of contact between α 3 and α 4 (Fig. 3c).

Interface 2 sequence conservation indicates a mechanism for dimer specificity within the DRP superfamily. Phylogenetic analysis of the hydrogen bonding partners within this region in dynamins, Drp1s and Dnm1, and Mx proteins allowed us to categorize three classes of residues: conserved in dynamins, Drp1/Dnm1, and Mx proteins (in

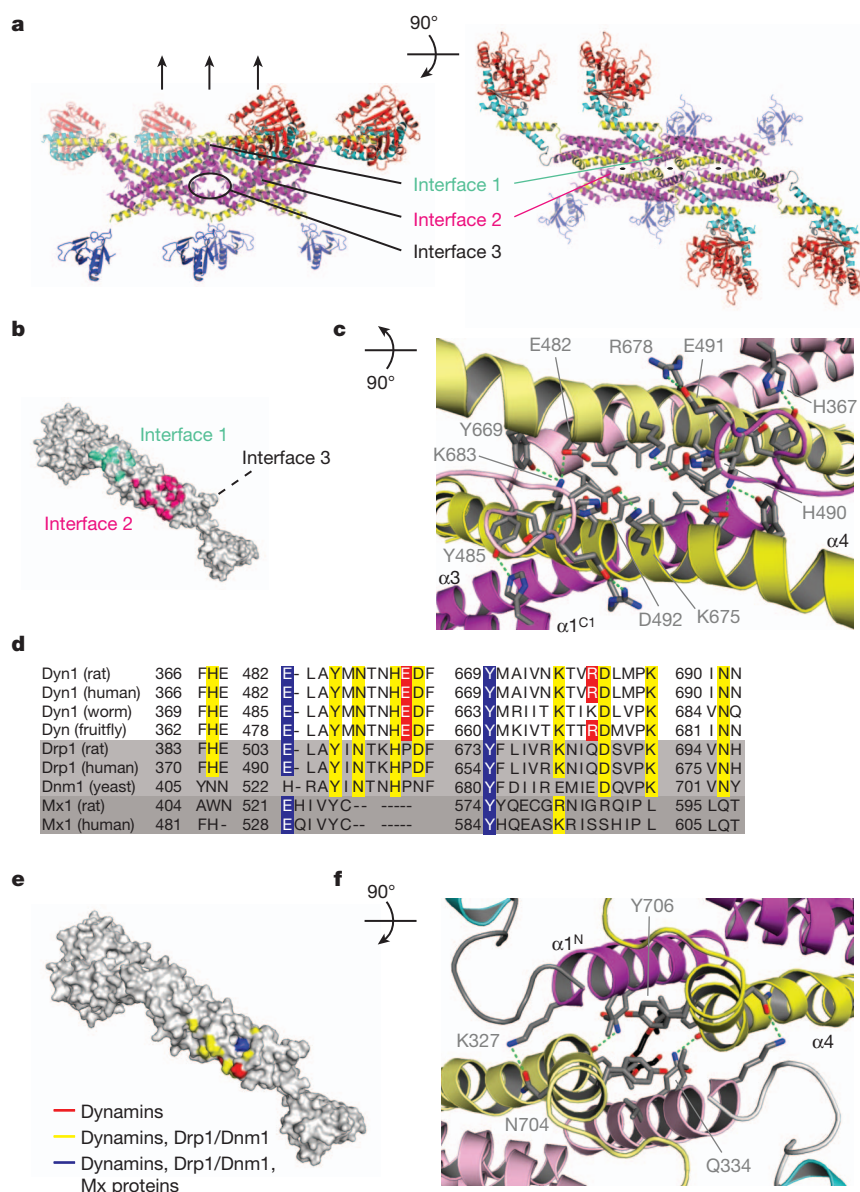


Figure 3 | Dyn1 G397D Δ PRD stalk interfaces mediate self-assembly.

a, Schematic diagram of four monomers showing interfaces in the crystal lattice. **b**, Surface representation showing the locations of interfaces 1 and 2. **c**, Detail of interface 2. Protomers are shown in lighter and darker hues. Green dotted lines are hydrogen bonds. **d**, Conservation of interface 2 residues involved in hydrogen bonding in dynamins, Drp1/Dnm1 and Mx proteins. Blue, conserved in dynamins, Drp1/Dnm1 and Mx proteins; yellow, conserved in dynamins and

Drps; red, conserved only within dynamins and Drp1/Dnm1 (in yellow), and conserved only in the dynamins (in red) (Fig. 3d). Plotting these classes onto a surface representation of the dynamin monomer (Fig. 3e) indicates that dimerization specificity is controlled by a spatial combinatorial code. As predicted from this code, hetero-oligomers consisting of dynamin 1 and dynamin 2 are observed³⁵. Several residues contributing to the apparent specificity are localized in the strongly conserved coil of dynamin, Drp1 and Dnm1 that follows $\alpha 3$.

Dynamine stalk interfaces drive self assembly

Interfaces 1 and 3 are also stalk-localized and mediate higher order assembly of dynamin dimers (Fig. 3a). Interface 1 is at the tips of interacting stalks, proximal to the GTPase domain and BSE (Fig. 3f) and is formed through interactions between protomer helices $\alpha 4$ and $\alpha 1^N$. The interface is capped by a stacking interaction between

Drps; red, conserved in dynamins. Alignment shows a subset of sequences used to determine the conservation. Sequences are identical to Supplementary Fig. 1, with the addition of *Rattus norvegicus* Mx1 (P18588.1) and *Homo sapiens* Mx1 (P20591.4). Fruitfly, *Drosophila melanogaster*; human, *Homo sapiens*; rat, *Rattus norvegicus*; worm, *Caenorhabditis elegans*; yeast, *Saccharomyces cerevisiae*. **e**, Surface representation of conservation data shown in (d). **f**, Interface 1. Density interpreted as a PEG400 molecule is shown in black.

opposing and flexible Y706 residues, conserved in dynamins and Drp1s. Interface 1 includes four hydrogen bonds (Fig. 3f, dashed green lines), with the remainder consisting of hydrophobic residues. The buried surface area of interface 1 is relatively small (624 Å²) and is likely to be dynamic to tolerate protomer interactions in a range of orientations.

Interface 3 is at the distal end of the stalk, where L2 loops from symmetry-related molecules are in close proximity (Supplementary Fig. 4b). The G397D mutation is within L2, which could not be traced due to poor density. The N-terminal region of L2 is also in close proximity to a symmetry-related L1, which features a highly conserved glycine, G346, that when mutated in Dnm1 (G385D) blocks assembly beyond the dimer¹⁶. We therefore predict that G397 and G346 are near one another in assembled dynamin. In addition, mutations at two dynamin arginine residues, R399 (in L2) and R361 (in $\alpha 1^{C1}$, following L1), and the corresponding interface 3 mutations in

MxA also inhibit assembly and stabilize dimeric forms^{18,27}. In contrast, disruption of MxA interface 1 yields a mixture of dimers and tetramers¹⁸. Thus, our data indicate that interface 3 is critical at an early step in the dynamin assembly pathway as its disruption stabilizes dimeric dynamin and allows linear filament formation. In contrast, interface 1 interactions probably function only to stabilize oligomerization.

From our structure, we propose that the dynamin dimer interface 2 is constitutive and relatively rigid. As compared to linear arrangement in the crystal, in helical assemblies we propose that the necessary rotational and translational shifts occur between adjacent dimers at interfaces 1 and 3 (Fig. 4a). Indeed, stalk dimers fit into the GMPPCP-bound helical electron microscopic reconstruction of dynamin²⁵ (Fig. 4b) possess a more tightly packed interface 3 and interface 1. These differences can be attributed to the disordered interface 3 in the crystal that is probably due to steric hindrance from the G397D mutation, and the dynamic nature of the hydrophobic interface 1.

Regulation of DRPs

The dynamin PH is essential for endocytosis and interacts with inositol phospholipids with low affinity^{12,36}. Centronuclear myopathy (CNM) disease mutations cluster at the C terminus of the PH α -helix (Supplementary Fig. 5a), underscoring its importance. They cause an increase in basal GTPase activity, without altering interactions with inositol phospholipids³³. In addition, SAXS analysis of an assembly-deficient dynamin indicates that the CNM mutants have a different conformation compared to wild type^{33,37}. In our lattice, three PH domains related by crystallographic symmetry lie close to interface 3 and the L1 loop (Fig. 2d), suggesting an interaction. Thus, the PH may serve to regulate access to this key multimerization interface to couple dynamin membrane interactions to dynamin assembly. Phosphoinositide binding by the PH variable loops and penetration of the membrane by variable loop 1¹⁴ could help to expose dynamin interface 3 and/or L1 and thus promote multimerization.

Alignment of the dynamin PH and corresponding sequences from Drp1s shows conservation of key residues (Supplementary Figs 1 and 5b), indicating that lipids may similarly regulate mammalian mitochondrial division. However, mammalian Drp1s lack most of the C-terminal PH α -helix, including residues mutated in CNM. The

details of Drp1 interface 3 regulation may therefore be different. Consistently, Drp1 splice variants have deletions in this region, pointing to a potential regulatory role. By contrast, *Danio rerio* Drp1 and fungal Dnm1 have an insert B of unknown structure in this region. Mitochondrial division DRPs can therefore be subdivided on this basis, which correlates with the divergence of their corresponding effectors, yeast Mdv1 and mammalian Mff^{38,39}.

Discussion

Our structure provides insight into the mechanism and regulation of dynamin assembly and into how the dynamin GTPase cycle is harnessed for function. Several observations point to a key role of the dynamin BSE in the formation of the GTPase–GTPase dimer interface observed in the GTPase–C_{GED} GDP•AlF₄[−] crystal structure¹⁰. This interface is likely to form from adjacent rungs of a dynamin helix and is critical to dynamin function as it mediates assembly-stimulated GTP hydrolysis. Comparison of the GTPase–GTPase dimer with our nucleotide-free structure indicates that in addition to expected differences, the BSE is flexible. Dynamin genetic data also support a role for the BSE in the regulation of GTPase–GTPase interface formation. Specifically, the dynamin switch 2 *shibire* ts2 mutation, G146S, which causes endocytic intermediates with ‘collared’ dynamin necks to accumulate⁴⁰, is suppressed by the *sushi* mutation A738T^{41,42}, located in the C_{GED} peptide, facing the hydrophobic groove (Fig. 5a). Insight into how dynamin undergoes conformational changes also comes from distantly related DRP structures (Supplementary Fig. 2). GTP-dependent GTPase domain dimerization is also observed for guanylate binding protein, indicating that this may be a common feature of the DRP superfamily²² (Supplementary Fig. 2c). In addition, two recent structures of the atlastin 1 cytosolic domain, thought to represent pre- and post-endoplasmic reticulum membrane fusion conformations^{23,24}, indicate that large changes occur in the position of the 3-helix bundle ‘middle domain’ relative to the GTPase domain (Supplementary Fig. 2b).

We propose that GTP binding and self-assembly promote dynamin GTPase–GTPase dimer formation via an opening of the BSE relative to the GTPase domain. In support, a modified version of our structure can be fit into the cryo-electron microscopic reconstruction of GMPPCP-bound dynamin with the BSE in a substantially more open

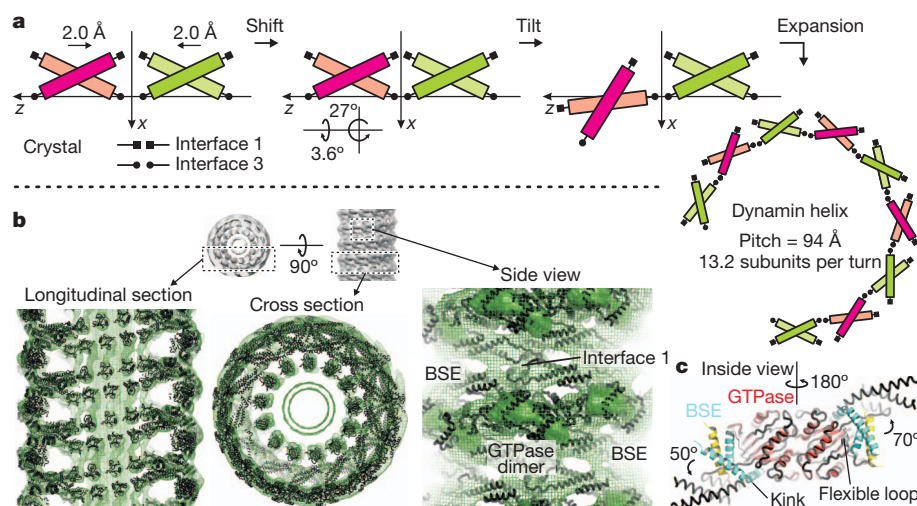


Figure 4 | Oligomerization of dynamin into helical structures. **a**, Dynamin helices derived from the linear arrangement in our crystal structure. Two stalk dimers (green and magenta) that engage in interface 1 and 3 are related by crystallographic translation. Experimentally determined helical parameters for dynamin assembled into helices in the GMPPCP-bound state²⁵ were matched by applying a small shift and tilt of one stalk dimer with respect to the other. **b**, Placement of oligomerized dynamin model into the electron microscopy

density map contoured at 1.2 σ . In side view: the fit of the GTPase domain as a GTPase–GTPase dimer with the BSE in open conformation to connect to interface 1 of the stalk helix (solid density is contoured at 3.6 σ). **c**, Observed conformational flexibility of the BSE. Model fitted into the helical reconstruction is shown as black superimposed ribbon on the crystal structure of the GTPase–C_{GED} fusion dimer (PDB accession no. 2X2E)¹⁰.

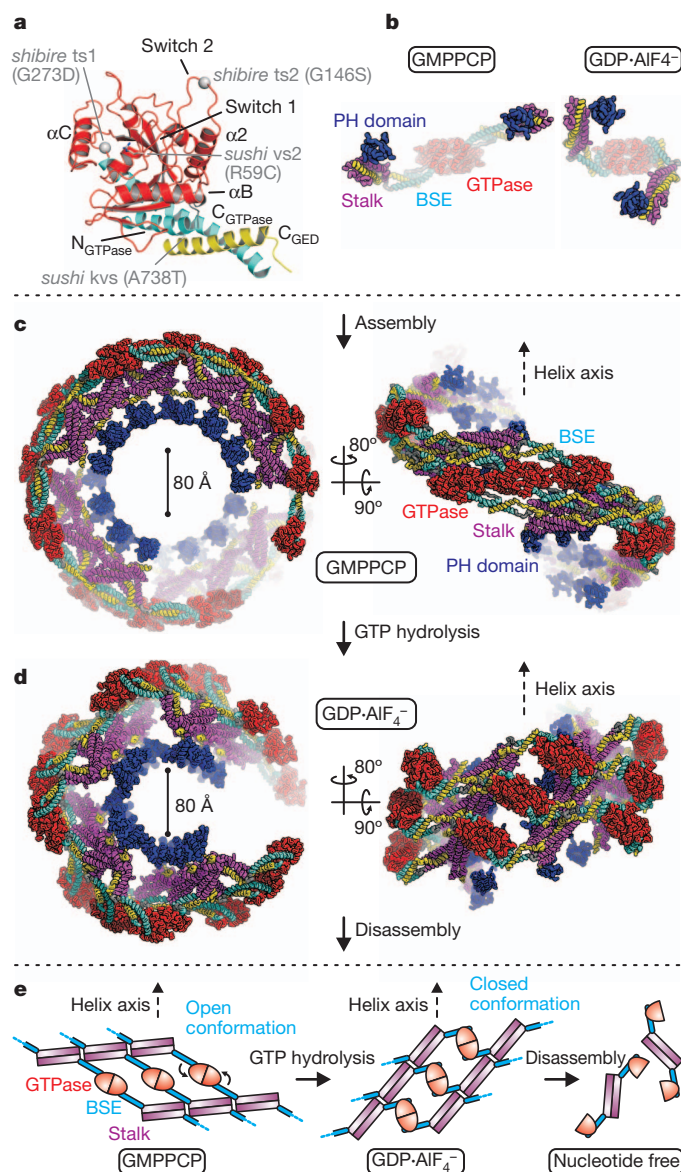


Figure 5 | Model for dynamin GTP cycle conformational changes.

a, Mapping of dynamin *shibire* and *sushi* mutations. **b**, Nucleotide-dependent dynamin conformations. The GTPase core domains (red) are in the same orientation. Left, GTP-bound state with open BSE conformation of dynamin as fitted into the GMPPCP-bound electron microscopic reconstruction shown in Fig. 4. Right, transition state of dynamin obtained by superposition of the BSE residues 291–312 and 727–743 of our structure on the corresponding residues of the GDP·AlF₄[−]-bound GTPase–C_{GED} fusion dimer (PDB accession no. 2X2E)¹⁰. Transition from open to closed BSE conformation results in movement of stalk domains. **c**, Model for Dyn1 ΔPRD GTP-bound helix. The BSE is opened to allow GTPase–GTPase dimer formation. **d**, GTP hydrolysis closes the BSE and adopts the conformation of the GDP·AlF₄[−]-bound transition state. This results in a substantial global constriction of the helical oligomeric assembly causing membrane deformation and scission. **e**, Schematic of how the proposed GTP hydrolysis triggered BSE conformational change is transmitted to oligomerized stalk domains.

conformation (Fig. 4b). We obtained a good fit of the GTPase–GTPase dimer domains and stalk interface 1 using constraints dictated by the helical fit of the oligomeric stalk and the strong GTPase domain-derived density (Fig. 4b). The BSE α -helices fit into the density stretching from the GTPase domain to the oligomeric stalk, indicating that a BSE flipping-out motion occurs at two hinge regions: the α -helical kink at P294 and in the loop connecting the N_{GTPase} and the GTPase domain, at P32 (Fig. 4c). Consistently, superposition of GDP·AlF₄[−]

GTPase–C_{GED} fusion GTPase–GTPase dimer partners indicates a smaller yet directionally equivalent opening of the BSE in one proto-monomer¹⁰. This proposed conformational change is feasible as there are relatively few contacts that hold the N_{GTPase} helix to the body of the GTPase domain in our structure. In addition, the BSE links the GTPase domain to the stalk and interface 1 via a short turn, making it well placed to transmit conformational alterations. As previously noted, comparison of the nucleotide-free and GMPPCP-bound electron microscopic reconstructions indicates that the GMPPCP-bound helix has a relatively smaller diameter¹⁷. In addition, the stalk density is more ‘kinked’ in the GMPPCP-bound form¹⁷. When we dock our structure, we observed that the bases of the stalks do not fit the density of the GMPPCP-bound helix. However, where the fit becomes poor there are strongly conserved prolines in α 2 and α 4 and a partially conserved proline in α 1^{C1} clustered in the stalk. We propose that these proline residues facilitate the formation of a kink at the stalk base in the GMPPCP-bound helix, which might allow interface 1 to form more fully in the assembled helix.

As predicted by our model, GTP hydrolysis would induce closure of the BSE via a transition state represented by the structure of the GTPase domain–BSE fusion in the presence of GDP·AlF₄[−], towards the conformation observed in our nucleotide-free structure (Fig. 5b). As GTP binding to dynamin is not rate limiting and GTP hydrolysis is stimulated by the formation of the GTPase–GTPase interface between adjacent rungs, it is likely that the BSE-dependent conformational change occurs in the context of short dynamin helical assemblies rather than within a helix consisting of many rungs^{8,32}. Short assemblies could also result in an approximate temporally coordinated conformational change (Fig. 5c–e). The conformational changes we propose would cause disruption to the assembled helix and the underlying membrane via local rung shifts. Interface 1 in our structure, which buries a relatively small surface area in the assembled stalk lattice, will be especially susceptible to change by inter-rung GTPase–GTPase dimer formation and its subsequent disruption by GTP hydrolysis, and these changes will be transmitted to interface 3 and the PH. The combined effects of curvature stress imposed by a short dynamin helical assembly coupled with PH insertion into the membrane is likely to destabilize the membrane and result in membrane fission¹⁴. Given the strong similarities between dynamins and other DRP family members, the structure of dynamin and our proposed model will serve to guide studies on the mechanisms of action of DRPs in diverse cellular functions.

METHODS SUMMARY

The conserved Dyn1 ΔPRD assembly-defective mutant G397D was identified by a cytological assay using the yeast mitochondrial fission DRP Dnm1–GFP. Dyn1 ΔPRD and Dyn1 G397D ΔPRD were expressed in *E. coli* and were purified as described in Methods. Light scattering, sucrose density gradient centrifugation, mass determination, continuous GTPase assays and electron microscopy were performed as described in Methods. Crystals were grown by microbatch from 3.2 μ l droplets containing 52.5 mM Tris/Cl pH 7.7, 175 mM NaCl, 32.5 mM NaNO₃, 20% v/v PEG 400, 0.97 mM β -mercaptoethanol and 31.9 μ M Dyn1 G397D ΔPRD and cryo-protected with ParetoN. Reflection data were collected at 100 K at Beamline 8.3.1 at the Advanced Light Source (Berkeley, California, USA) at a wavelength of 0.9488 Å. Data collection and processing are described in Methods. The structure was determined by molecular replacement, using known structures of the nucleotide-free rat dynamin 1 GTPase domain (PDB accession no. 2AKA chain B), the human dynamin 1 PH domain (PDB accession no. 2DYN chain B) and a truncated form of the human MxA stalk (PDB accession no. 3LJB chain B) as sequential search models. Structure refinement is described in Methods. The stalk interface 2 dimer was fit into the previously described cryo-electron microscopic reconstruction of GMPPCP-bound dynamin by applying tilt and twist to the sequential dimers from the linear filaments observed in the crystal to match the helical parameters described for the reconstruction. The previously described GTPase domain dimer formed in the presence of GDP·AlF₄[−] was subsequently fit into the density. Connection of the GTPase domain to the stalk required a conformational rearrangement of the BSE, which was independently fit into the visible density. The fit was subjected to rigid body refinement as described in Methods.

Received 10 May; accepted 12 August 2011.

Published online 18 September 2011.

1. Praefcke, G. J. & McMahon, H. T. The dynamin superfamily: universal membrane tubulation and fission molecules? *Nature Rev. Mol. Cell Biol.* **5**, 133–147 (2004).
2. Hoppins, S., Lackner, L. & Nunnari, J. The machines that divide and fuse mitochondria. *Annu. Rev. Biochem.* **76**, 751–780 (2007).
3. Marks, B. *et al.* GTPase activity of dynamin and resulting conformational change are essential for endocytosis. *Nature* **410**, 231–235 (2001).
4. Stowell, M. H. B., Marks, B., Wigge, P. & McMahon, H. T. Nucleotide-dependent conformational changes in dynamin: evidence for a mechanochemical molecular spring. *Nature Cell Biol.* **1**, 27–32 (1999).
5. Mears, J. A. *et al.* Conformational changes in Dnm1 support a contractile mechanism for mitochondrial fission. *Nature Struct. Mol. Biol.* **18**, 20–26 (2011).
6. Warnock, D. E., Hinshaw, J. E. & Schmid, S. L. Dynamin self-assembly stimulates its GTPase activity. *J. Biol. Chem.* **271**, 22310–22314 (1996).
7. Danino, D., Moon, K. H. & Hinshaw, J. E. Rapid constriction of lipid bilayers by the mechanochemical enzyme dynamin. *J. Struct. Biol.* **147**, 259–267 (2004).
8. Pucadyil, T. J. & Schmid, S. L. Real-time visualization of dynamin-catalyzed membrane fission and vesicle release. *Cell* **135**, 1263–1275 (2008).
9. Chappie, J. S. *et al.* An intramolecular signaling element that modulates dynamin function *in vitro* and *in vivo*. *Mol. Biol. Cell* **20**, 3561–3571 (2009).
10. Chappie, J. S., Acharya, S., Leonard, M., Schmid, S. L. & Dyda, F. G domain dimerization controls dynamin's assembly-stimulated GTPase activity. *Nature* **465**, 435–440 (2010).
11. Timm, D. *et al.* Crystal structure of the pleckstrin homology domain from dynamin. *Nature Struct. Biol.* **1**, 782–788 (1994).
12. Achiriloaie, M., Barylko, B. & Albanesi, J. P. Essential role of the dynamin pleckstrin homology domain in receptor-mediated endocytosis. *Mol. Cell Biol.* **19**, 1410–1415 (1999).
13. Ferguson, K. M., Lemmon, M. A., Schlessinger, J. & Sigler, P. B. Crystal structure at 2.2 Å resolution of the pleckstrin homology domain from human dynamin. *Cell* **79**, 199–209 (1994).
14. Ramachandran, R. *et al.* Membrane insertion of the pleckstrin homology domain variable loop 1 is critical for dynamin-catalyzed vesicle scission. *Mol. Biol. Cell* **20**, 4630–4639 (2009).
15. Hinshaw, J. E. & Schmid, S. L. Dynamin self-assembles into rings suggesting a mechanism for coated vesicle budding. *Nature* **374**, 190–192 (1995).
16. Ingberman, E. *et al.* Dnm1 forms spirals that are structurally tailored to fit mitochondria. *J. Cell Biol.* **170**, 1021–1027 (2005).
17. Mears, J. A., Ray, P. & Hinshaw, J. E. A corkscrew model for dynamin constriction. *Structure* **15**, 1190–1202 (2007).
18. Gao, S. *et al.* Structural basis of oligomerization in the stalk region of dynamin-like MxA. *Nature* **465**, 502–506 (2010).
19. Low, H. H., Sachse, C., Amos, L. A. & Lowe, J. Structure of a bacterial dynamin-like protein lipid tube provides a mechanism for assembly and membrane curving. *Cell* **139**, 1342–1352 (2009).
20. Low, H. H. & Lowe, J. A bacterial dynamin-like protein. *Nature* **444**, 766–769 (2006).
21. Prakash, B., Praefcke, G. J., Renault, L., Wittinghofer, A. & Herrmann, C. Structure of human guanylate-binding protein 1 representing a unique class of GTP-binding proteins. *Nature* **403**, 567–571 (2000).
22. Ghosh, A., Praefcke, G. J., Renault, L., Wittinghofer, A. & Herrmann, C. How guanylate-binding proteins achieve assembly-stimulated processive cleavage of GTP to GMP. *Nature* **440**, 101–104 (2006).
23. Bian, X. *et al.* Structures of the atlastin GTPase provide insight into homotypic fusion of endoplasmic reticulum membranes. *Proc. Natl Acad. Sci. USA* **108**, 3976–3981 (2011).
24. Byrnes, L. J. & Sondermann, H. Structural basis for the nucleotide-dependent dimerization of the large G protein atlastin-1/SPG3A. *Proc. Natl Acad. Sci. USA* **108**, 2216–2221 (2011).
25. Zhang, P. & Hinshaw, J. E. Three-dimensional reconstruction of dynamin in the constricted state. *Nature Cell Biol.* **3**, 922–926 (2001).
26. Siegel, L. M. & Monty, K. J. Determination of molecular weights and frictional ratios of proteins in impure systems by use of gel filtration and density gradient centrifugation. Application to crude preparations of sulfite and hydroxylamine reductases. *Biochim. Biophys. Acta* **112**, 346–362 (1966).
27. Ramachandran, R. *et al.* The dynamin middle domain is critical for tetramerization and higher-order self-assembly. *EMBO J.* **26**, 559–566 (2007).
28. Gao, S. *et al.* Structural basis of oligomerization in the stalk region of dynamin-like MxA. *Nature* **465**, 502–506 (2010).
29. Song, B. D., Yarar, D. & Schmid, S. L. An assembly-incompetent mutant establishes a requirement for dynamin self-assembly in clathrin-mediated endocytosis *in vivo*. *Mol. Biol. Cell* **15**, 2243–2252 (2004).
30. Reubold, T. F. *et al.* Crystal structure of the GTPase domain of rat dynamin 1. *Proc. Natl Acad. Sci. USA* **102**, 13093–13098 (2005).
31. Sweitzer, S. M. & Hinshaw, J. E. Dynamin undergoes a GTP-dependent conformational change causing vesiculation. *Cell* **93**, 1021–1029 (1998).
32. Bashkurov, P. V. *et al.* GTPase cycle of dynamin is coupled to membrane squeeze and release, leading to spontaneous fission. *Cell* **135**, 1276–1286 (2008).
33. Kenniston, J. A. & Lemmon, M. A. Dynamin GTPase regulation is altered by PH domain mutations found in centronuclear myopathy patients. *EMBO J.* **29**, 3054–3067 (2010).
34. Niemann, H. H., Knetsch, M. L., Scherer, A., Manstein, D. J. & Kull, F. J. Crystal structure of a dynamin GTPase domain in both nucleotide-free and GDP-bound forms. *EMBO J.* **20**, 5813–5821 (2001).
35. Barylko, B. *et al.* The proline/arginine-rich domain is a major determinant of dynamin self-activation. *Biochemistry* **49**, 10592–10594 (2010).
36. Vallis, Y., Wigge, P., Marks, B., Evans, P. R. & McMahon, H. T. Importance of the pleckstrin homology domain of dynamin in clathrin-mediated endocytosis. *Curr. Biol.* **9**, 257–263 (1999).
37. Durieux, A. C., Prudhon, B., Guicheney, P. & Bitoun, M. Dynamin 2 and human diseases. *J. Mol. Med.* **88**, 339–350 (2010).
38. Lackner, L. L., Horner, J. S. & Nunnari, J. Mechanistic analysis of a dynamin effector. *Science* **325**, 874–877 (2009).
39. Gandre-Babbe, S. & van der Bliek, A. M. The novel tail-anchored membrane protein Mff controls mitochondrial and peroxisomal fission in mammalian cells. *Mol. Biol. Cell* **19**, 2402–2412 (2008).
40. Kosaka, T. & Ikeda, K. Possible temperature-dependent blockage of synaptic vesicle recycling induced by a single gene mutation in *Drosophila*. *J. Neurobiol.* **14**, 207–225 (1983).
41. Ramaswami, M., Rao, S., van der Bliek, A., Kelly, R. B. & Krishnan, K. S. Genetic studies on dynamin function in *Drosophila*. *J. Neurogenet.* **9**, 73–87 (1993).
42. Narayanan, R., Leonard, M., Song, B. D., Schmid, S. L. & Ramaswami, M. An internal GAP domain negatively regulates presynaptic dynamin *in vivo*: a two-step model for dynamin function. *J. Cell Biol.* **169**, 117–126 (2005).
43. Ingberman, E. & Nunnari, J. A continuous, regenerative coupled GTPase assay for dynamin-related proteins. *Methods Enzymol.* **404**, 611–619 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors would like to express thanks to I. Stokes-Rees for assistance with the Wide Space Molecular Replacement, J. Holton for advice and assistance with data collection and H. McMahon for the kind gift of the dynamin clone. We would also like to thank J. Al-Bassam, J. Chappie, A. McCoy, S. Harrison, D. Owen, A. Fisher, E. Baldwin, L. Lackner, G. Adamson, N. Varlakhanova and members of the Nunnari lab for extensive discussions. S.J. is a Damon Runyon Cancer Research Foundation Fellow supported by the Howard Hughes Medical Institute (DRG-2004-09). This work was supported by an American Heart Innovator Award and NIH grants (R01GM062942S1 and R01GM097432) to J.N. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the US Department of Energy under Contract No. DE-AC02-05CH11231.

Author Contributions M.G.J.F. purified, biochemically characterized and crystallized the Dyn1 G397D ΔPRD. M.G.J.F. collected X-ray diffraction data and M.G.J.F. and S.J. solved the structure. M.G.J.F. and J.N. designed experiments and M.G.J.F., S.J. and J.N. interpreted data and prepared the manuscript.

Author Information Atomic coordinates and structure factors for Dyn1 G397D ΔPRD have been deposited with the Protein Data Bank under accession code 3ZVR. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.N. (jmnunnari@ucdavis.edu).

Gravitational redshift of galaxies in clusters as predicted by general relativity

Radosław Wojtak¹, Steen H. Hansen¹ & Jens Hjorth¹

The theoretical framework of cosmology is mainly defined by gravity, of which general relativity is the current model. Recent tests of general relativity within the Lambda Cold Dark Matter (Λ CDM) model have found a concordance between predictions and the observations of the growth rate and clustering of the cosmic web^{1,2}. General relativity has not hitherto been tested on cosmological scales independently of the assumptions of the Λ CDM model. Here we report an observation of the gravitational redshift of light coming from galaxies in clusters at the 99 per cent confidence level, based on archival data³. Our measurement agrees with the predictions of general relativity and its modification created to explain cosmic acceleration without the need for dark energy (the $f(R)$ theory⁴), but is inconsistent with alternative models designed to avoid the presence of dark matter^{5,6}.

According to the theory of general relativity⁷, light emitted from galaxies moving in the gravitational potential well of galaxy clusters is expected to be redshifted proportionally to the difference in gravitational potential Φ between the clusters and an observer, that is, $z_{\text{gr}} = \Delta\Phi/c^2$, where c is the velocity of light in vacuum. For typical cluster masses of $\sim 10^{14} M_{\odot}$, where M_{\odot} is the Sun's mass, the gravitational redshift is estimated to be $^{8-10} cz_{\text{gr}} \approx 10 \text{ km s}^{-1}$, which is around two orders of magnitude smaller than the Doppler shift owing to the random motions of galaxies in clusters. The method of disentangling the kinematic Doppler effect from gravitational redshift relies on the fact that the former gives rise to a symmetric broadening of the observed velocity distribution, whereas the latter shifts its centroid. A critical factor in detecting such a velocity shift is the number of galaxies with spectroscopically measured velocities and the number of galaxy clusters. Both should be sufficiently high to reduce the error due to the Doppler width of the velocity distribution and to eliminate the sensitivity to irregularities in cluster structure, such as substructures and asphericity.

The data are compiled from the SDSS³ Data Release 7 and the associated Gaussian Mixture Brightest Cluster Galaxy catalogue¹¹ containing the positions and redshifts of galaxy clusters identified in the survey. The cluster sample is richness-limited with a threshold corresponding to a cluster mass of $10^{14} M_{\odot}$. The mean, 5th percentile and 95th percentile values of the cluster richness¹¹ are 16, 8, and 86 and correspond to cluster masses of around $2 \times 10^{14} M_{\odot}$, $10^{14} M_{\odot}$ and $10^{15} M_{\odot}$, respectively. The typical number of spectroscopic redshifts per cluster (within a 6-megaparsec (Mpc) aperture and a velocity range of $\pm 4,000 \text{ km s}^{-1}$ around the mean cluster velocities) varies from 10 for low-richness clusters to 140 for the richest ones.

Figure 1 shows the histograms of galaxy velocities calculated in four bins of the projected cluster-centric distance centred at 0.6, 1.6, 3.3 and 5.2 Mpc. The cluster centres and redshifts were approximated by the coordinates and redshifts of the brightest cluster galaxies. The observed velocity distributions consist of two clearly distinct parts: a quasi-flat distribution of galaxies not belonging to the clusters (observed due to projection effect) and a quasi-Gaussian component associated with galaxies gravitationally bound to the clusters¹². The latter is expected to reveal the signature of gravitational redshift in terms of a systematic shift of its velocity centroid. Analysis of mock

kinematic data generated from cosmological simulations shows that the number of redshifts and clusters is sufficient to reduce all expected sources of noise—such as substructures, cluster asphericity, and non-negligible offset between the brightest cluster galaxies and cluster centres¹³ (both in the position on the sky and redshift space)—and to allow for detection of gravitational redshift at a confidence level of nearly 3σ (see Supplementary Information).

We search for gravitational redshift by measuring the mean velocity Δ of the quasi-Gaussian component of the observed velocity distribution. We carry out a Monte Carlo Markov Chain analysis of the data using a two-component model for the velocity distribution which includes contributions from both the cluster and non-cluster galaxies (Supplementary Information). Constraints on the mean velocity are obtained by marginalizing the likelihood function over the set of nuisance parameters defining the shape of both components of the velocity distribution. The best-fitting models of the velocity distributions are shown in Fig. 1 and the resulting measurements of the mean velocity as a function of the projected cluster-centric distance R are presented in Fig. 2. The obtained mean velocity is negative at all radii with a clear tendency to decline with increasing radius. The negative values arise from the fact that the rest frames of the clusters are defined by the observed velocities of the central galaxies. This choice of the reference frame implies that the gravitational redshift manifests itself

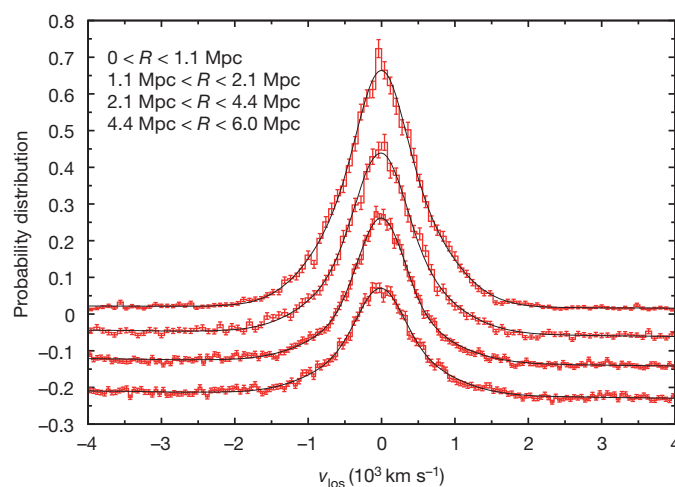


Figure 1 | Velocity distributions of galaxies combined from 7,800 SDSS galaxy clusters. The line-of-sight velocity (v_{los}) distributions are plotted in four bins of the projected cluster-centric distances R (indicated in the upper left corner for the distributions from top to bottom), and offset vertically by an arbitrary amount for presentation purposes. Red lines present the histograms of the observed galaxy velocities in the cluster rest frame and black solid lines show the best-fitting models. The model assumes a linear contribution from the galaxies that do not belong to the cluster and a quasi-Gaussian contribution from the cluster members (see Supplementary Information for more details). The cluster rest frames and centres are defined by the redshifts and the positions of the brightest cluster galaxies. The error bars represent Poisson noise.

¹Dark Cosmology Centre, Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, DK-2100 Copenhagen Ø, Denmark.

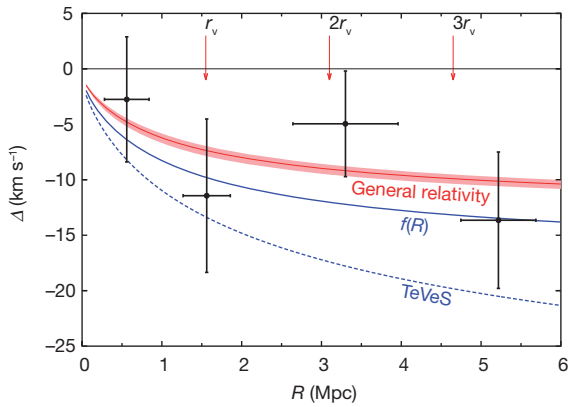


Figure 2 | Constraints on gravitational redshift in galaxy clusters. The effect manifests itself as a blueshift Δ of the velocity distributions of cluster galaxies in the rest frame of their brightest cluster galaxies. Velocity shifts were estimated as the mean velocity of a quasi-Gaussian component of the observed velocity distributions (see Fig. 1). The error bars represent the range of Δ parameter containing 68% of the marginal probability and the dispersion of the projected radii in a given bin. The blueshift (black points) varies with the projected radius R and its value at large radii indicates the mean gravitational potential depth in galaxy clusters. The red profile represents theoretical predictions of general relativity calculated on the basis of the mean cluster gravitational potential inferred from fitting the velocity dispersion profile using the assumptions of the most reliable anisotropic model of galaxy orbits (see Supplementary Information for more details). Its width shows the range of Δ containing 68% of the marginal probability. The blue solid and dashed lines show the profiles corresponding to two modifications of standard gravity: $f(R)$ theory⁴ and the TeVeS model^{5,6}. Both profiles were calculated on the basis of the corresponding modified gravitational potentials (see Supplementary Information for more details). The prediction for $f(R)$ represents the case that maximizes the deviation from the gravitational acceleration in standard gravity on the scales of galaxy clusters. Assuming isotropic orbits in fitting the velocity dispersion profile lowers the mean gravitational depth of the clusters by 20%. The resulting profiles of gravitational redshift for general relativity and $f(R)$ theory are still consistent with the data and the discrepancy between prediction of TeVeS and the measurements remains nearly the same. The arrows show characteristic scales related to the mean radius r_v of the virialized parts of the clusters.

as a blueshift¹⁰ (negative mean velocity) varying with the projected cluster-centric distance from 0 at the cluster centre to $-|\Phi(0)|/c$ at large projected radii R .

The detection of gravitational redshift is significant at the 99% confidence level. The integrated signal within the 6-Mpc aperture amounts to $\Delta = -7.7 \pm 3.0 \text{ km s}^{-1}$, which is consistent with the gravitational potential depths of simulated galaxy clusters of $\Delta = -(5\text{--}10) \text{ km s}^{-1}$ (ref. 10). A more quantitative comparison with theoretical predictions requires explicit information about the mean gravitational potential profile and the distribution of cluster masses in the sample. We make use of the velocity dispersion profile of the composite cluster to constrain both functions. Then we calculate the gravitational redshift in terms of the mean velocity Δ by convolving the individual profiles of the clusters with their mass distribution (Supplementary Information). The resulting profile (red profile in Fig. 2; see also discussion on the effect of the anisotropy of galaxy orbits in Supplementary Information) is fully consistent with the gravitational redshift inferred from the velocity distributions. The fact that the same gravitational potential underlies galaxy motions and gravitational redshift of photons in clusters provides observational evidence of the equivalence principle on the scale of galaxy clusters.

We confront the obtained constraints on gravitational redshift with the predictions of alternative theories of gravity. We consider two popular models of gravity, the tensor–vector–scalar (TeVeS) theory^{5,6} and the $f(R)$ model⁴, designed to alleviate the problem of dark matter or to recover the expansion history of the Universe, respectively. Theoretical profiles of gravitational redshift are calculated using the

relations between the generalized gravitational potentials of these models and the Newtonian potential (Supplementary Information). The Newtonian potential is inferred from the observed velocity dispersion profile of the composite cluster using the assumptions of the most reliable anisotropic model of galaxy orbits (see Supplementary Information for more details), and constitutes the reference basis for the calculations. For TeVeS we assume that the total masses of galaxy clusters make up 80% of those recovered under assumption of the Newtonian gravity. This factor lowers the ratio of dynamical-to-baryonic mass in galaxy clusters to the value resulting from fitting Modified Newtonian Dynamics⁵ (to which TeVeS is a relativistic generalization) to cluster data¹⁴. The resulting profile of gravitational redshift does not match the data, deviating from the observations at the 95% confidence level (the blue dashed line in Fig. 2). This discrepancy increases with projected radius and is mostly caused by a logarithmic divergence of the scalar field in the regime of small accelerations, that is, $g < a_0$ and $a_0 \approx 10^{-10} \text{ m s}^{-2}$, which is responsible for a $1/r$ modification of the gravitational acceleration. This result points to a critical problem for the TeVeS (or Modified Newtonian Dynamics) model in recovering the true gravitational potential at large distances around the cluster centres. Considering the $f(R)$ model, we choose the least favourable set of free parameters maximizing the departure from Newtonian gravitational acceleration¹⁵. Despite this choice, the resulting profile of gravitational redshift is consistent with the data (the blue solid line in Fig. 2).

The obtained constraints on gravity are consistent with recent tests verifying the concordance between gravity, the cosmological model and observations of the large-scale structure of the Universe^{1,2}. An important advantage of using the gravitational redshift effect is that this method does not depend on cosmology (see also Supplementary Information) allowing to probe gravity in a direct way. In particular, this implies that the discrepancy between TeVeS theory and the observations² is unlikely

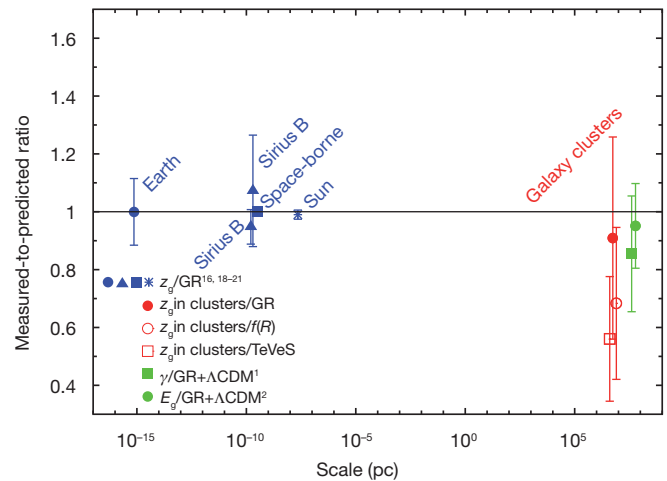


Figure 3 | The measured-to-predicted ratio of the gravitational redshift. The figure shows the results of different observations or experiments as a function of the spatial scale of the gravitational potential well. Blue and red symbols refer to detections of gravitational redshift z_g in: a ground-based experiment¹⁶ (blue circle), observations of the Sirius B white dwarf^{19,20} (blue triangles), a space-based experiment²¹ (blue square), an observation of the Sun¹⁸ (blue star), and analysis of the cluster data reported in this work (red circle). All measurements are divided by the predictions of general relativity (filled symbols). Results obtained for galaxy clusters are also compared with the predictions of $f(R)$ theory and TeVeS model (red open symbols). As a measure of gravitational redshift in galaxy clusters we used the signal integrated within the aperture of 6 Mpc. The green square and circle show the measurement of the rate of growth of cosmic structure γ (ref. 1) and the probe of gravity E_g , which combines the properties of galaxy–galaxy lensing, galaxy clustering and galaxy velocities². Both results are compared with the prediction of general relativity with a standard Λ CDM cosmological model. All error bars represent standard deviations. The relative accuracy of the measurement from the space-borne experiment²¹ is beyond the resolution of the plot and amounts to 10^{-4} .

to be a consequence of a specific choice of cosmological parameters, but indeed points to the inadequacy of this model to describe the Universe on very large scales.

Our results complement a series of experiments and observations aimed at confirmation of the predicted gravitational redshift on different scales of the Universe. Figure 3 shows a summary of the detections in terms of the relative accuracy of the measurements as a function of the scale of the gravitational potential well. The positions of data points vary from 20 m, for the first ground-based experiment^{16,17}, to the 1–10-Mpc scale for galaxy clusters. On the scale of the solar radius we plot the measurement of gravitational redshift for the Sun¹⁸, and on a scale two orders of magnitude smaller we plot constraints from the observations of the Sirius B white dwarf^{19,20} and a space-borne hydrogen maser²¹. These results make gravitational redshift the only effect predicted by general relativity that has been confirmed on spatial scales spanning 22 orders of magnitude. To study this effect in more detail requires a larger redshift sample and therefore will be possible with the advent of the next-generation redshift surveys, such as the EUCLID satellite.

Received 26 May; accepted 11 August 2011.

- Rapetti, D. *et al.* The observed growth of massive galaxy clusters—III. Testing general relativity on cosmological scales. *Mon. Not. R. Astron. Soc.* **406**, 1796–1804 (2010).
- Reyes, R. *et al.* Confirmation of general relativity on large scales from weak lensing and galaxy velocities. *Nature* **464**, 256–258 (2010).
- Abazajian, K. N. *et al.* The Seventh Data Release of the Sloan Digital Sky Survey. *Astrophys. J. Suppl. Ser.* **182**, 543–558 (2009).
- Carroll, S. *et al.* Is cosmic speed-up due to new gravitational physics? *Phys. Rev. D* **70**, 043528 (2004).
- Milgrom, M. A modification of the Newtonian dynamics—implications for galaxies. *Astrophys. J.* **270**, 371–389 (1983).
- Bekenstein, J. D. Relativistic gravitation theory for the modified Newtonian dynamics paradigm. *Phys. Rev. D* **70**, 083509 (2004).
- Einstein, A. Die Grundlage der allgemeinen Relativitätstheorie. *Astron. Nachr.* **354**, 769–822 (1916).
- Capri, A. Gravitational redshift in galaxy clusters. *Astron. Astrophys.* **301**, 6–10 (1995).
- Broadhurst, T. & Scannapieco, E. Detecting the gravitational redshift of cluster gas. *Astrophys. J.* **533**, L93–L97 (2000).
- Kim, Y.-R. & Croft, R. A. C. Gravitational redshifts in simulated galaxy clusters. *Astrophys. J.* **607**, 164–174 (2004).
- Hao, J. *et al.* A GMBCG galaxy cluster catalog of 55,424 rich clusters from SDSS DR7. *Astrophys. J. Suppl. Ser.* **191**, 254–274 (2010).
- Wojtak, R. *et al.* Interloper treatment in dynamical modelling of galaxy clusters. *Astron. Astrophys.* **466**, 437–449 (2007).
- Skibba, R. A. *et al.* Are brightest halo galaxies central galaxies? *Mon. Not. R. Astron. Soc.* **410**, 417–431 (2011).
- Pointecouteau, E. & Silk, J. New constraints on modified Newtonian dynamics from galaxy clusters. *Mon. Not. R. Astron. Soc.* **364**, 654–658 (2005).
- Schmidt, F. Dynamical masses in modified gravity. *Phys. Rev. D* **81**, 103002 (2010).
- Pound, R. V. & Rebka, G. A. Gravitational redshift in nuclear resonance. *Phys. Rev. Lett.* **3**, 439–441 (1959).
- Pound, R. V. & Snider, J. L. Effect of gravity on nuclear resonance. *Phys. Rev. Lett.* **13**, 539–540 (1964).
- Lopresto, J. C., Schrader, C. & Pierce, A. K. Solar gravitational redshift from the infrared oxygen triplet. *Astrophys. J.* **376**, 757–760 (1991).
- Greenstein, J. L., Oke, J. B. & Shipman, H. L. Effective temperature, radius, and gravitational redshift of Sirius B. *Astrophys. J.* **169**, 563–566 (1971).
- Barstow, M. A. *et al.* Hubble Space Telescope spectroscopy of the Balmer lines in Sirius B. *Mon. Not. R. Astron. Soc.* **362**, 1134–1142 (2005).
- Vessot, R. F. C. *et al.* Test of relativistic gravitation with a space-borne hydrogen maser. *Phys. Rev. Lett.* **45**, 2081–2084 (1980).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The Dark Cosmology Centre is funded by the Danish National Research Foundation. R.W. wishes to thank D. Rapetti, G. Mamon and S. Gottlöber for discussions and suggestions. The mock catalogues of galaxy clusters have been obtained from a simulation performed at the Altix of the LRZ Garching.

Author Contributions R.W. was responsible for the analysis of the velocity distributions and the velocity dispersion profile, predictions for the models of modified gravity, and drafting the manuscript. S.H.H. was responsible for the comparison with the constraints on gravitational redshift on different scales, and writing and commenting on the paper. J.H. conceived the idea of the measurement, and wrote and commented on the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.W. (wojtak@dark-cosmology.dk).

The excitation of solar-like oscillations in a δ Sct star by efficient envelope convection

V. Antoci¹, G. Handler^{1,2}, T. L. Campante^{3,4}, A. O. Thygesen^{4,5}, A. Moya⁶, T. Kallinger^{1,7,8}, D. Stello⁹, A. Grigahcène³, H. Kjeldsen⁴, T. R. Bedding⁹, T. Lüftinger¹, J. Christensen-Dalsgaard⁴, G. Catanzaro¹⁰, A. Frasca¹⁰, P. De Cat¹¹, K. Uytterhoeven^{12,13,14,15}, H. Bruntt⁴, G. Houdek¹, D. W. Kurtz¹⁶, P. Lenz², A. Kaiser¹, J. Van Cleve¹⁷, C. Allen¹⁸ & B. D. Clarke¹⁷

Delta Scuti (δ Sct)¹ stars are opacity-driven pulsators with masses of $1.5\text{--}2.5M_{\odot}$, their pulsations resulting from the varying ionization of helium. In less massive stars² such as the Sun, convection transports mass and energy through the outer 30 per cent of the star and excites a rich spectrum of resonant acoustic modes. Based on the solar example, with no firm theoretical basis, models predict that the convective envelope in δ Sct stars extends only about 1 per cent of the radius³, but with sufficient energy to excite solar-like oscillations^{4,5}. This was not observed before the Kepler mission⁶, so the presence of a convective envelope in the models has been questioned. Here we report the detection of solar-like oscillations in the δ Sct star HD 187547, implying that surface convection operates efficiently in stars about twice as massive as the Sun, as the *ad hoc* models predicted.

Thirty days of continuous observations of HD 187547 (KIC 7548479) by the Kepler mission with a cadence of 1 min led to its identification as a δ Sct pulsator (Fig. 1a, b). In contrast to the non-uniformly distributed signals at low frequencies, the observed regularly spaced peaks at high frequencies (Fig. 1c) suggest that we also observe high-radial-order overtones as expected for stochastically excited solar-like oscillations. For such oscillations the observed comb-like frequency structure (with the large frequency separation $\Delta\nu$ indicating the frequency separation between consecutive radial overtones of like degree) is the result of mainly radial and dipolar pulsation modes, whereas for δ Sct stars it is not clear which modes are excited to observable amplitudes. The strikingly broadened structures observed only at high frequencies (Figs 1f and 2b, c) suggest that each is due either to single damped and stochastically re-excited oscillations or to very close unresolved frequencies of coherent oscillations.

Here we use spectroscopic observations to derive an effective temperature $T_{\text{eff}} = 7,500 \pm 250$ K, a surface gravity of $\log g = 3.90 \pm 0.25$ dex (c.g.s.) and a projected rotational velocity of $v \sin i = 10.3 \pm 2.3$ km s⁻¹ (see Supplementary Information for details). We identify HD 187547 as an Am star from chemical element abundance analysis, which is consistent with the observed low $v \sin i$ typical for these stars. Am stars are stars of spectral type A showing atmospheric underabundance when compared with the Sun in the chemical elements Sc and Ca, and an overabundance of Ba, Sr and Y (ref. 7). We compute a photospheric metallicity (all elements except H and He) of $Z = 0.017$, which is larger than the solar value of $Z = 0.0134$ (ref. 8).

About two-thirds of Am stars are primary components of spectroscopic binary systems⁹. The Am phenomenon is connected to slow

rotation, which is not common in A type stars. Binarity is believed to act as a braking mechanism slowing down the rotation and allowing spectral peculiarities to occur as a result of element diffusion¹⁰. Pulsating Am stars still represent a challenge to theory, because He is expected to settle gravitationally and should only partly be present in the He II ionization zone where the δ Sct pulsations are excited. In other words, theoretical models predict that the hottest and youngest A-type stars should not pulsate¹⁰, which is in contradiction with recent observations¹¹. As the stars evolve, their convective envelopes deepen and efficiently mix the stellar matter, erasing the observed chemical peculiarities in the atmospheres, allowing the opacity mechanism to drive pulsation in the He II ionization zone. Using the observed solar-like oscillations reported here, the depth of the convective envelope can be derived (hence the mixing length), probing the diffusion of He and heavy elements in this star. This will contribute significantly to revising the interaction between pulsation and diffusion in models of Am stars.

Seven radial velocity measurements of HD 187547, spread over 153 days, give no evidence for a short-period binary system. In addition, the absence of any detectable contribution by a potential close companion to the spectrum implies a considerably less luminous star of spectral type G or later. The expected amplitudes and frequency of maximum oscillation power for such a star are inconsistent with the observations, leading to the conclusion that the signal observed in Fig. 1c cannot originate from a companion. The observed amplitude spectrum of HD 187547 is not affected by a background star because the fraction of light in the aperture from neighbouring stars is only 1.5%. Other chemically peculiar pulsating stars situated, as the δ Sct stars, in the classical instability strip in the Hertzsprung–Russell diagram¹² are the rapidly oscillating Ap stars. Their high-radial-order pulsation modes are triggered by the opacity mechanism acting in the hydrogen ionization zone, often showing equidistant multiplets in the frequency spectrum as a result of the alignment of the pulsation axes with strong magnetic fields¹³. The strong magnetic fields as seen in rapidly oscillating Ap stars are, however, not observed in Am stars¹⁴. We therefore exclude the possibility that HD 187547 is a hybrid of a δ Sct and a rapidly oscillating Ap star.

In Fig. 3 we show an échelle diagram comparing the observed frequencies with a model of a star similar to HD 187547, demonstrating again the clear structures separated by $\Delta\nu$ at high frequencies and the non-structured distribution at lower frequencies. For the high-frequency modes we derive a mean large frequency separation $\Delta\nu$ of 40.5 ± 0.6 μ Hz. Using the empirical relation¹⁵

¹Institute of Astronomy, University of Vienna, Türkenschanzstraße 18, A-1180 Vienna, Austria. ²Copernicus Astronomical Center, Bartycza 18, 00-716 Warsaw, Poland. ³Centro de Astrofísica, Departamento de Física e Astronomia-Faculdade de Ciências, Universidade do Porto, Rua das Estrelas, 4150-762 Porto, Portugal. ⁴Department of Physics and Astronomy, Aarhus University, Ny Munkegade 120, DK-8000 Aarhus C, Denmark. ⁵Nordic Optical Telescope, Apartado 474, E-38700 Santa Cruz de La Palma, Santa Cruz de Tenerife, Spain. ⁶Departamento de Astrofísica, Centro de Astrobiología, Instituto Nacional de Técnica Aeroespacial – Consejo Superior de Investigaciones Científicas, PO Box 78, E-28691, Villanueva de la Cañada, Madrid, Spain. ⁷Department of Physics and Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver, British Columbia V6T 1Z1, Canada. ⁸Instituut voor Sterrenkunde, University of Leuven, Celestijnenlaan 200D, 3001 Leuven, Belgium. ⁹Sydney Institute for Astronomy (SIfA), School of Physics, University of Sydney, New South Wales 2006, Australia. ¹⁰Istituto Nazionale di Astrofisica – Osservatorio Astrofisico di Catania, via S. Sofia 78, I-95123 Catania, Italy. ¹¹Royal Observatory of Belgium, Ringlaan 3, B-1180 Brussels, Belgium. ¹²Laboratoire AIM, CEA/DSM-CNRS Université Paris Diderot, Commissariat à l'Energie Atomique et aux Energies Alternatives, Institut de Recherche sur Les Lois Fondamentales, Service d'Astrophysique, Centre de Saclay, F-91191, Gif-sur-Yvette, France. ¹³Kiepenheuer-Institut für Sonnenphysik, Schöneckstrasse 6, 79104 Freiburg, Germany. ¹⁴Instituto de Astrofísica de Canarias, 38200 La Laguna, Tenerife, Spain. ¹⁵Departamento de Astrofísica, Universidad de La Laguna, 38205 La Laguna, Tenerife, Spain. ¹⁶Jeremiah Horrocks Institute, University of Central Lancashire, Preston PR1 2HE, UK. ¹⁷SETI Institute/NASA Ames Research Center, Moffett Field, California 94035, USA. ¹⁸Orbital Sciences Corporation/NASA Ames Research Center, Moffett Field, California 94035, USA.

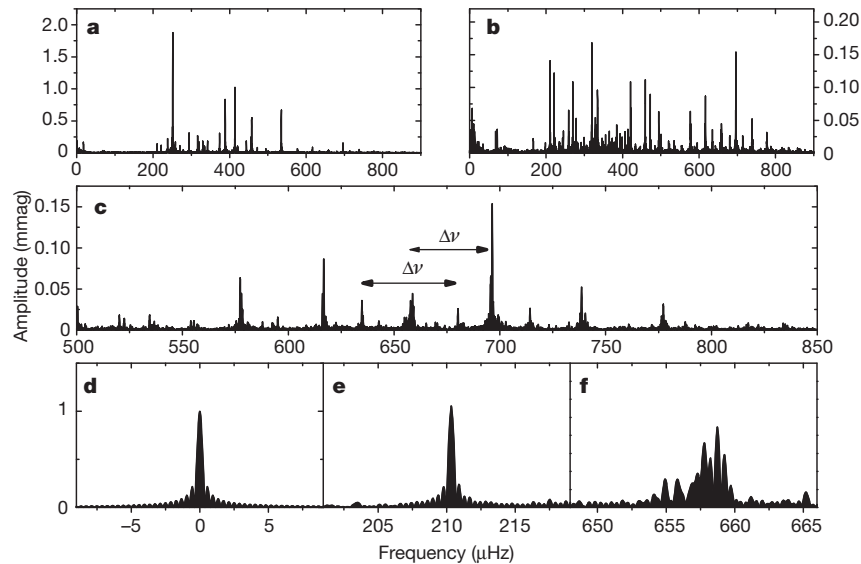


Figure 1 | Fourier amplitude spectra of the Kepler light curve of HD 187547. **a**, Fourier spectrum covering the entire frequency range in which significant signals were observed with a dominant frequency at 251 μHz and an amplitude of 2 mmag, typical for a δ Sct star. **b**, The multimode oscillations of HD 187547 are shown by subtracting 16 sinusoids corresponding to the most prominent oscillations, revealing a large number of additional significant frequencies. **c**, The region between 500 and 850 μHz shows a clear pattern of roughly equally spaced peaks, which we interpret as high-order consecutive radial overtones. The comb-like structure expected for high-order radial

overtones is clearly visible. The broadened peaks suggest damped/re-excited solar-like oscillations. The black arrows denoted $\Delta\nu$ indicate the large separation between consecutive radial and dipole modes. **d**, Spectral window. The shape of the window function is defined by the length and sampling of the data set. Any coherent signal will have the same profile. **e**, Example for one of the modes driven by the opacity mechanism in HD 187547. **f**, A supposed solar-like oscillation mode observed in HD 187547, displaying a broadened structure suggestive of a short mode lifetime.

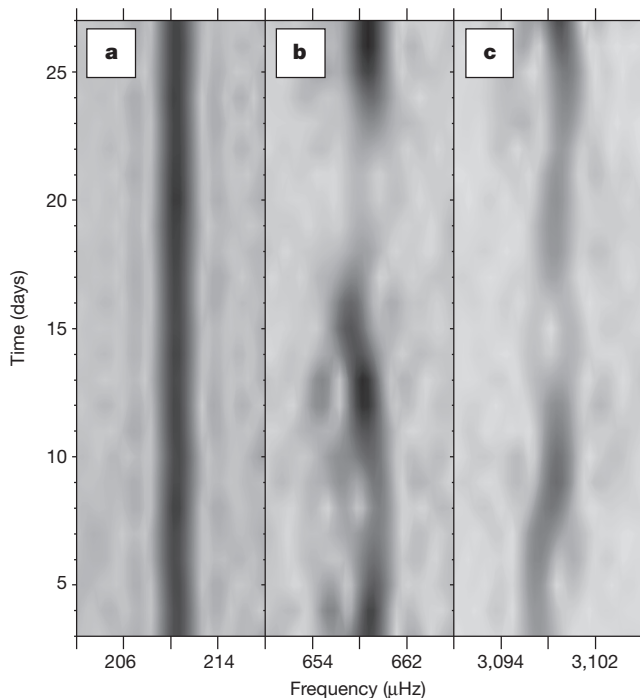


Figure 2 | Time-Fourier spectrum. Here we again highlight the difference in temporal variability between the modes interpreted as stochastic modes and the coherent, opacity-driven peaks at low frequencies. The time-fourier spectrum was computed with a running filter of full-width at half-maximum = 5 days, comparable to the mean mode lifetime. **a**, An opacity-driven mode (the same as in Fig. 1e) showing temporal stability in the δ Sct frequency region. **b**, Stochastic mode observed in HD 187547, showing an erratic behaviour as expected for solar-like oscillations (the same as in Fig. 1f). **c**, For comparison, a stochastic oscillation mode observed in the Sun. The solar data were obtained from the SOHO VIRGO instrument. The data set has the same length and sampling as for HD 187547; that is, 30 days and 1 min, respectively. Further details of frequency analyses and tests on artificial data sets (Supplementary Fig. 1) to verify our interpretation are in the Supplementary Information.

$\Delta\nu = (0.263 \pm 0.009) \mu\text{Hz}(\nu_{\text{max}} \mu\text{Hz}^{-1})^{0.772 \pm 0.005}$ we obtain a frequency of maximum power $\nu_{\text{max}} = 682^{+41}_{-43} \mu\text{Hz}$. This is in very good agreement with the highest-amplitude mode in the supposed stochastic frequency region at 696 μHz . The possibility that what we observe is $0.5\Delta\nu$ in the frequency spectrum is ruled out because this would require a ν_{max} at about 1,673 μHz , where no signal is observed. We can also exclude the observation of $2\Delta\nu$ because that would place ν_{max} at 277 μHz , close to the dominant opacity-driven mode at 251 μHz .

The amplitudes of solar-like oscillations are determined by the interaction between driving and damping defined by different physical processes², such as modulation of the turbulent momentum and heat fluxes by pulsation. The exact contribution to driving and damping by each of these processes is still not well understood, resulting in uncertainties in the predictions of the stochastically excited mode amplitudes¹⁶, particularly in hotter stars^{2,3} in which the convective envelopes are shallow. We expect the mixing length, the amplitudes and mode lifetimes to constrain the anisotropy of the convective velocity field, parameters that all semi-analytical convection models rely on¹⁷.

For HD 187547 we measure a peak-amplitude per radial mode¹⁸ for the assumed stochastic signal of 56 ± 2 p.p.m., which after bolometric correction¹⁹ results in 67 ± 3 p.p.m. (see Supplementary Information for details). From the empirical scaling relation²⁰ and using a bolometric solar peak-amplitude of 3.6 p.p.m. (ref. 21) we obtain a predicted peak amplitude of $A = 14 \pm 9$ p.p.m. The mean mode lifetimes are measured²² as 5.7 ± 0.8 days. Empirical relations predict a mode lifetime for a star with $T_{\text{eff}} = 7,500 \pm 250$ K of the order of one day²³ or shorter²⁴, which is not in agreement with what we measure for HD 187547. However, these scaling relations (for amplitude and mode lifetimes) are based on few observed stars, and none of them is calibrated in the temperature domain of our target, for which the physical conditions in the convection zone are expected to be very different. Furthermore, given that HD 187547 is metal overabundant in comparison with the Sun, the observed amplitude is expected to be higher^{3,25} than predicted from simple scaling, which is indeed the case. The power of a mode is directly proportional to the mode lifetime provided that the energy supply rate over the mode inertia is constant²⁶, which further supports

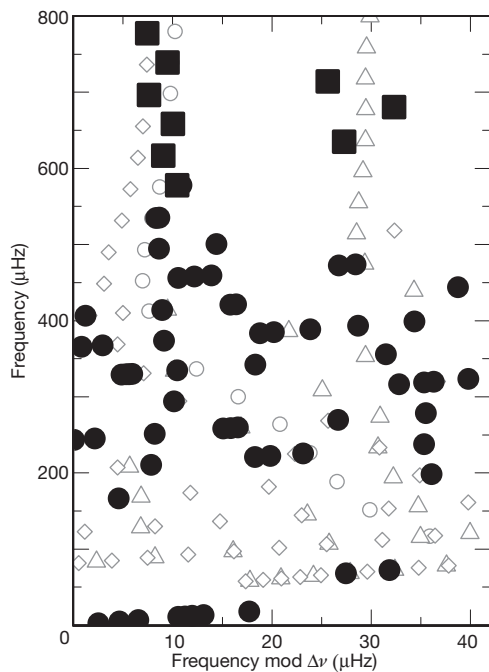


Figure 3 | Échelle diagram of HD 187547. Here we plot 69 extracted frequencies as a function of frequency modulo the large separation ($\Delta\nu = 40.5 \mu\text{Hz}$). Frequencies equally spaced by $\Delta\nu$ will form vertical ridges in the échelle diagram. To guide the eye, we show theoretically predicted frequencies of pulsation modes for a $1.85M_{\odot}$ stellar model. Ridges of $l = 0$ modes are represented by open circles, $l = 1$ by open triangles and $l = 2$ by open diamonds. Detailed modelling of the star is beyond the scope of this paper. The supposed solar-like modes (filled squares) between 500 and 870 μHz show clear ridges, as expected for high-order acoustic oscillations similar to what is observed in solar-like stars. The lower frequencies that we attribute to δ Sct pulsation (filled circles) excited by the opacity mechanism show no obvious regular patterns.

the higher amplitude because the observed mode lifetimes are also longer than expected. An additional factor that is not considered in any scaling relations is the chemical peculiarity of our target. In summary, these factors make HD 187547 an intriguing case for further theoretical analyses of stochastic oscillations and the potential interaction with the opacity mechanism in δ Sct stars.

The amplitude distribution for stochastic pulsation can be described as a Rayleigh distribution, provided that the examined time series are much shorter than the mode lifetimes. The relation between the mean amplitude $\langle A \rangle$ and its standard deviation $\sigma(A)$ can then be written as²⁷ $(4/\pi - 1)^{0.5} \langle A \rangle \approx 0.52 \langle A \rangle$. This is not valid for opacity-driven pulsation. For HD 187547 we therefore expect to obtain two different regimes of the ratio $\sigma(A)/\langle A \rangle$ for the two groups of oscillation modes (Supplementary Fig. 2). Indeed, we see that the δ Sct frequencies have a lower value of $\sigma(A)/\langle A \rangle$ than the supposed solar-like modes, giving further evidence for the stochastic nature of the latter (see Supplementary Information for details).

We cannot strictly exclude the possibility that the signals between 578 and 868 μHz are due to unresolved modes of pulsation excited by the opacity mechanism, because high-radial-order acoustic modes can also be observed in hot δ Sct stars. Nevertheless, as shown in Fig. 1 this would imply that δ Sct pulsation covers the region between 205 and 870 μHz continuously. According to current theory, the opacity mechanism acting in the He II ionization zone cannot excite modes spanning 16 radial orders for a star with parameters like those of HD 187547 (ref. 28). Further support for the discovery of solar-like oscillations comes from spectroscopic observations that also indicate the presence of convective motions in the atmospheres of A and Am stars²⁹. In addition, signatures of granulation noise in δ Sct stars have been

reported from photometric measurements³. Opacity-driven pulsations are also observed in more massive stars ($8\text{--}16M_{\odot}$), known as β Cephei stars (in this case the opacity mechanism acts in the ionization region of the iron-group elements). The unexpected detection of solar-like oscillations in such a star³⁰ (with a mass of $10M_{\odot}$) suggests that both types of pulsation, opacity-driven and stochastically excited, can coexist and can have overlapping frequency domains. The similar timescales of the different oscillation types imply a possible interaction between the two mechanisms.

Received 7 April; accepted 26 July 2011.

Published online 14 September; corrected 29 September 2011 (see full-text HTML version for details).

1. Breger, M. in *Delta Scuti and Related Stars* (ASP Conf. Ser. Vol. 210) (eds Breger, M. & Montgomery, M.) 3–42 (Astronomical Society of the Pacific, 2000).
2. Chaplin, W. J. et al. Ensemble asteroseismology of solar-type stars with the NASA Kepler mission. *Science* **332**, 213–216 (2011).
3. Kallinger, T. & Matthews, J. M. Evidence for granulation in early A-type stars. *Astrophys. J.* **711**, L35–L39 (2010).
4. Houdek, G., Balmforth, N. J., Christensen-Dalsgaard, J. & Gough, D. O. Amplitudes of stochastically excited oscillations in main-sequence stars. *Astron. Astrophys.* **351**, 582–596 (1999).
5. Samadi, R., Goupil, M.-J. & Houdek, G. Solar-like oscillations in delta Scuti stars. *Astron. Astrophys.* **395**, 563–571 (2002).
6. Koch, D. G. et al. Kepler mission design, realized photometric performance, and early science. *Astrophys. J.* **713**, L79–L86 (2010).
7. Preston, G. W. The chemically peculiar stars of the upper main sequence. *Annu. Rev. Astron. Astrophys.* **12**, 257–277 (1974).
8. Asplund, M., Grevesse, N., Sauval, A. J. & Scott, P. The chemical composition of the Sun. *Annu. Rev. Astron. Astrophys.* **47**, 481–522 (2009).
9. Carquillat, J.-M. & Prieur, J.-L. Contribution to the search for binaries among Am stars. VIII. New spectroscopic orbits of eight systems and statistical study of a sample of 91 Am stars. *Mon. Not. R. Astron. Soc.* **380**, 1064–1078 (2007).
10. Turcotte, S., Richer, J., Michaud, G. & Christensen-Dalsgaard, J. The effect of diffusion on pulsations of stars on the upper main sequence— δ Scuti and metallic A stars. *Astron. Astrophys.* **360**, 603–616 (2000).
11. Balona, L. et al. Kepler observations of Am stars. *Mon. Not. R. Astron. Soc.* **414**, 792–800 (2011).
12. Handler, G. Confirmation of simultaneous p and g mode excitation in HD 8801 and γ Peg from time-resolved multicolour photometry of six candidate ‘hybrid’ pulsators. *Mon. Not. R. Astron. Soc.* **398**, 1339–1351 (2009).
13. Kurtz, D. W. Rapidly oscillating AP stars. *Mon. Not. R. Astron. Soc.* **200**, 807–859 (1982).
14. Aurière, M. et al. No detection of large-scale magnetic fields at the surfaces of Am and HgMn stars. *Astron. Astrophys.* **523**, A40, doi:10.1051/0004-6361/201014848 (2010).
15. Stello, D., Chaplin, W. J., Basu, S., Elsworth, Y. & Bedding, T. R. The relation between $\Delta\nu$ and ν_{max} for solar-like oscillations. *Mon. Not. R. Astron. Soc.* **400**, L80–L84 (2009).
16. Houdek, G. Solar-type variables. *AIP Conf. Proc.* **1170**, 519–530 (2009).
17. Samadi, R., Belkacem, K., Goupil, M.-J., Kupka, F. & Dupret, M.-A. Solarlike oscillation amplitudes and line-widths as a probe for turbulent convection in stars. *Proc. Int. Astron. Un.* **239**, 349–357 (2007).
18. Kjeldsen, H. et al. The amplitude of solar oscillations using stellar techniques. *Astrophys. J.* **682**, 1370–1375 (2008).
19. Ballot, J., Barban, C. & van’t Veer-Menneret, C. Visibilities and bolometric corrections for stellar oscillation modes observed by Kepler. *Astron. Astrophys.* **531**, A124, doi:10.1051/0004-6361/201016230 (2011).
20. Kjeldsen, H. & Bedding, T. R. Amplitudes of solar-like oscillations: a new scaling relation. *Astron. Astrophys.* **529**, L8, doi:10.1051/0004-6361/201116789 (2011).
21. Michel, E. et al. Intrinsic photometric characterisation of stellar oscillations and granulation. *Astron. Astrophys.* **495**, 979–987 (2009).
22. Gruberbauer, M., Kallinger, T., Weiss, W. W. & Guenther, D. B. On the detection of Lorentzian profiles in a power spectrum: a Bayesian approach using ignorance priors. *Astron. Astrophys.* **506**, 1043–1053 (2009).
23. Chaplin, W. J., Houdek, G., Karoff, G., Elsworth, Y. & New, R. Mode lifetimes of stellar oscillations. Implications for asteroseismology. *Astron. Astrophys.* **500**, L21–L24 (2009).
24. Baudin, F. et al. Amplitudes and lifetimes of solar-like oscillations observed by CoRoT. Red-giant versus main-sequence stars. *Astron. Astrophys.* **529**, A84, doi:10.1051/0004-6361/201014037 (2011).
25. Samadi, R., Ludwig, H.-G., Belkacem, K., Goupil, M. J. & Dupret, M.-A. The CoRoT target HD 49933. I. Effect of the metal abundance on the mode excitation rates. *Astron. Astrophys.* **509**, A15, doi:10.1051/0004-6361/200911867 (2010).
26. Chaplin, W. J. et al. On model predictions of the power spectral density of radial solar p modes. *Mon. Not. R. Astron. Soc.* **360**, 859–868 (2005).
27. Chang, H.-Y. & Gough, D. O. On the power distribution of solar p modes. *Solar Phys.* **181**, 251–263 (1998).
28. Pamyatnykh, A. A. in *Delta Scuti and Related Stars* (ASP Conf. Ser. Vol. 210) (eds Breger, M. & Montgomery, M.) 215–246 (Astronomical Society of the Pacific, 2000).

29. Landstreet, J. D. *et al.* Atmospheric velocity fields in tepid main sequence stars. *Astron. Astrophys.* **503**, 973–984 (2009).
30. Belkacem, K. *et al.* Solar-like oscillations in a massive star. *Science* **324**, 1540–1542 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the entire Kepler team, without whom these results would not be possible. V.A., G.Ha. and G.Ho. were supported by the Austrian Fonds zur Förderung der wissenschaftlichen Forschung. V.A. also thanks L. Fossati for his help. A.M. acknowledges the funding of AstroMadrid, who was also supported by Spanish grants. T.R.B. and D.S. acknowledge support from the Australian Research Council. T.L. was supported by the Austrian Agency for International Cooperation in Education and Research. K.U. acknowledges financial support from the Deutsche Forschungsgemeinschaft. Funding for this Discovery mission is provided by NASA's Science Mission Directorate.

Author Contributions V.A. discovered the star among the Kepler targets, analysed it and found the solar-like oscillations (as a part of her PhD thesis), did spectroscopic analyses, frequency analyses, the test on the stochastic nature of the signal, and interpretations, and wrote the paper. G.Ha. had the idea for this project and supervised V.A., helped with analyses, interpretations and writing the paper. T.L.C. contributed to the analyses of the stochastic modes and also to the test on the stochastic nature of the signal. A.O.T. observed the target spectroscopically at Nordic Optical Telescope, identified the star as

an Am star and did spectroscopic analyses. A.M. contributed to the statistical test on the nature of the stochastic signal. T.K. helped interpretations, data analyses and writing the paper, and delivered the mode lifetimes. D.S. helped with data analyses, writing the paper and made Fig. 2. A.G. helped with theoretical interpretations and the time-Fourier analyses. T.R.B. helped with interpretations and writing the paper. H.K. contributed to the analyses, also by supervising V.A., and is a member of the Kepler Asteroseismic Investigation Steering Committee. J.C.-D. helped with the theoretical support, writing the paper and is a member of the Kepler Asteroseismic Investigation Steering Committee. T.L. confirmed the Am identification, excluding the Ap character of the star. G.C., A.F. and A.K. did spectroscopic analyses. P.D.C. was Principal Investigator and observer for the spectroscopic data from McDonald observatory. K.U. was Co-Investigator of the McDonald data and coordinated the ground-based observations. H.B. was Principal Investigator for the observations with the NARVAL spectrograph and did spectroscopic analyses. G.Ho. and P.L. helped with theoretical interpretations and writing the paper. D.W.K. helped with the Am classification and writing the paper and is leader of the delta Scuti working group of the Kepler Asteroseismic Science Consortium. J.V.C., C.A. and B.D.C. are part of the Kepler team and were involved in designing and operating the satellite. All co-authors contributed to discussions and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to V.A. (victoria.antoci@univie.ac.at).

High-fidelity projective read-out of a solid-state spin quantum register

Lucio Robledo^{1*}, Lilian Childress^{2*}, Hannes Bernien^{1*}, Bas Hensen¹, Paul F. A. Alkemade¹ & Ronald Hanson¹

Initialization and read-out of coupled quantum systems are essential ingredients for the implementation of quantum algorithms^{1,2}. Single-shot read-out of the state of a multi-qubit (multi-qubit) register would allow direct investigation of quantum correlations (entanglement), and would give access to further key resources such as quantum error correction and deterministic quantum teleportation¹. Although spins in solids are attractive candidates for scalable quantum information processing, their single-shot detection has been achieved only for isolated qubits^{3–6}. Here we demonstrate the preparation and measurement of a multi-spin quantum register in a low-temperature solid-state system by implementing resonant optical excitation techniques originally developed in atomic physics. We achieve high-fidelity read-out of the electronic spin associated with a single nitrogen-vacancy centre in diamond, and use this read-out to project up to three nearby nuclear spin qubits onto a well-defined state⁷. Conversely, we can distinguish the state of the nuclear spins in a single shot by mapping it onto, and subsequently measuring, the electronic spin^{5,8}. Finally, we show compatibility with qubit control: we demonstrate initialization, coherent manipulation and single-shot read-out in a single experiment on a two-qubit register, using techniques suitable for extension to larger registers. These results pave the way for a test of Bell's inequalities on solid-state spins and the implementation of measurement-based quantum information protocols.

The electronic spin of the nitrogen-vacancy centre (NV) in diamond constitutes an exceptional solid-state system for investigating quantum phenomena, combining excellent spin coherence^{9–12} with a robust optical interface^{13–16}. Furthermore, the host nitrogen nuclear spin (typically ¹⁴N, with nuclear spin $I = 1$) and nearby isotopic impurity ¹³C nuclei ($I = 1/2$) have hyperfine interactions with the NV's electronic spin ($S = 1$), allowing development of few-spin quantum registers that have been suggested as building blocks for quantum repeaters¹⁷, cluster state computation¹⁸ and distributed quantum computing¹⁹. All of these applications require high-fidelity preparation, manipulation and measurement of multiple spins. There have been significant advances in coherent control over few-spin systems in diamond^{20,21}, but no method exists for the simultaneous preparation^{22,23} and single-shot read-out⁵ of multi-spin registers, which impedes progress towards multi-qubit protocols. Here we remove this obstacle by exploiting resonant excitation techniques, as pioneered in atomic physics^{24,25}, in microstructured diamond devices that allow high photon collection efficiency (Fig. 1a). These new methods enable us to initialize multiple nuclear spin qubits and to perform single-shot read-out of a few-qubit register, clearing the way towards implementation of quantum algorithms with solid-state spins.

Our preparation and read-out techniques rely on resonant excitation of spin-selective optical transitions of the NV, which can be spectrally resolved at low temperatures²⁶. We use the E_x and A_1 transitions in our experiments (Fig. 1b): A_1 connects the ground states with spin projection $m_S = \pm 1$ to an excited state with a primarily

$m_S = \pm 1$ character, whereas E_x connects states with $m_S = 0$. A typical spectrum of NV A, one of the two NVs we study, is shown in Fig. 1c (see Supplementary Information for NV B). Under resonant excitation of a single transition, the fluorescence decays with time owing to a slight spin mixing within the excited states that induces shelving into the other spin state (Fig. 1d). This optical pumping mechanism allows high-fidelity spin state initialization^{24,27}: from the data in Fig. 1d, we estimate a preparation error into the $m_S = 0$ ground state of $0.3 \pm 0.1\%$, which is a drastic reduction of the $11 \pm 3\%$ preparation

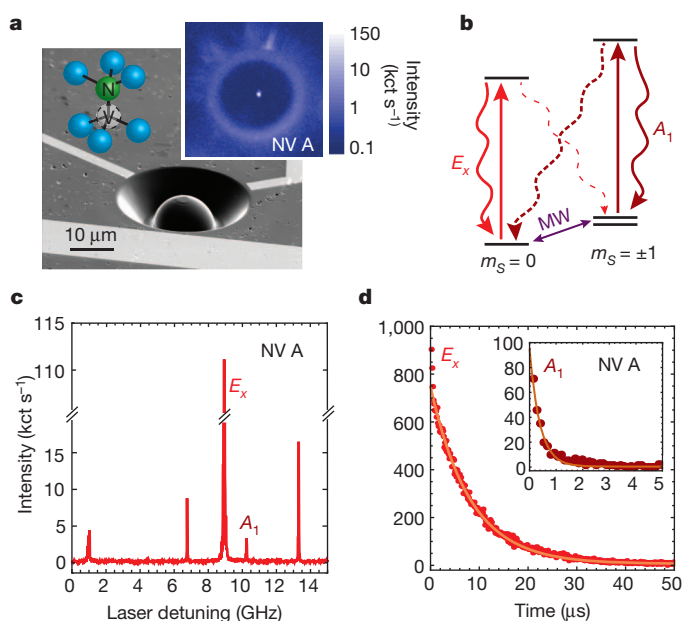


Figure 1 | Resonant excitation and electronic spin preparation of a nitrogen-vacancy centre. **a**, Scanning electron microscope image of a solid immersion lens representative of those used in the experiments (for details, see Supplementary Information). The overlaid sketch shows the substitutional nitrogen and the adjacent vacancy that form the NV. Inset, scanning confocal microscope image of NV A (logarithmic colour scale). kct, 1,000 counts. **b**, Energy levels used to prepare and read out the NV's electronic spin ($S = 1$ in the ground and optically excited states); transitions are labelled according to the symmetry of their excited states. Dashed lines indicate spin-non-conserving decay paths. MW, microwave transition. **c**, Photoluminescence excitation spectrum of NV A; frequency is given relative to 470.443 THz. **d**, Fluorescence time trace of NV A, initially prepared in $m_S = 0$ (E_x excitation, 4.8-nW power) and $m_S = \pm 1$ (A_1 excitation, 7.4-nW power; inset), with a saturation power $P_{\text{sat}} \approx 6$ nW. Spin flips in the excitation cycle lead to nearly exponential decay of fluorescence, with fitted spin-flip times of $1/\gamma_0 = 8.1 \pm 0.1$ μs for E_x and 0.39 ± 0.01 μs for A_1 , and initial respective intensities of 740 ± 5 and 95 ± 2 kct s⁻¹, giving a lower limit of $99.7 \pm 0.1\%$ to the $m_S = 0$ preparation fidelity and one of $99.2 \pm 0.1\%$ to the $m_S = \pm 1$ preparation fidelity. The low initial intensity for A_1 is associated with a fast intersystem crossing to metastable singlet states (Supplementary Information).

¹Kavli Institute of Nanoscience Delft, Delft University of Technology, PO Box 5046, 2600 GA Delft, The Netherlands. ²Department of Physics and Astronomy, Bates College, 44 Campus Avenue, Lewiston, Maine 04240, USA.

*These authors contributed equally to this work.

error observed with conventional off-resonant initialization (Supplementary Information).

Spin-dependent resonant excitation also allows single-shot electronic spin read-out: the presence or absence of fluorescence under E_x excitation reveals the spin state. By working with low-strain NVs at low temperature ($T = 8.6$ K), we suppress spin mixing^{26,28} and phonon-induced transitions²⁹ within the excited states, extending the spin relaxation time under E_x excitation to several microseconds. Together with a high collection efficiency due to the use of solid immersion lenses³⁰ fabricated around pre-selected, low-strain NVs, and efficient rejection of resonant excitation from the measured phonon-sideband emission, this highly spin-preserving transition allows the detection of several photons before the spin flips.

We demonstrate single-shot read-out by initializing the electronic spin to have $m_S = 0$ or $m_S = \pm 1$, followed by resonant excitation of the E_x read-out transition for $t_{ro} = 100$ μ s (Fig. 2a). The resulting histograms of the number of detected photons are given in Fig. 2b. As expected, for $m_S = \pm 1$ we observe negligible excitation, with a 98.3%

probability of not measuring any photons during the probe interval. By stark contrast, after initializing the spin to have $m_S = 0$ we detect an average of $\langle n_0 \rangle = 6.4$ photons per shot. We assign the state $m_S = 0$ to detection of one or more photons, and $m_S = \pm 1$ to the detection of no photons. After truncating our integration window to the optimal duration of 40 μ s, we find an average fidelity of

$$F_{\text{avg}} = \frac{1}{2} (F_{m_S=0} + F_{m_S=\pm 1}) = 93.2 \pm 0.5\%$$

where F_{m_S} is the probability of obtaining the measurement outcome m_S after optical pumping into m_S . To verify that these measurement outcomes indeed correspond to the electronic spin states, we use single-shot read-out to observe spin Rabi oscillations and microwave-induced quantum jumps²⁵ (Fig. 2d,e).

Whereas the full read-out optically pumps the spin, shorter read-out durations can be non-destructive, albeit at lower fidelity. By optimizing integration windows, we obtain a fidelity of $83.4 \pm 0.5\%$ for each of two successive read-out segments (Fig. 2c). Correlations

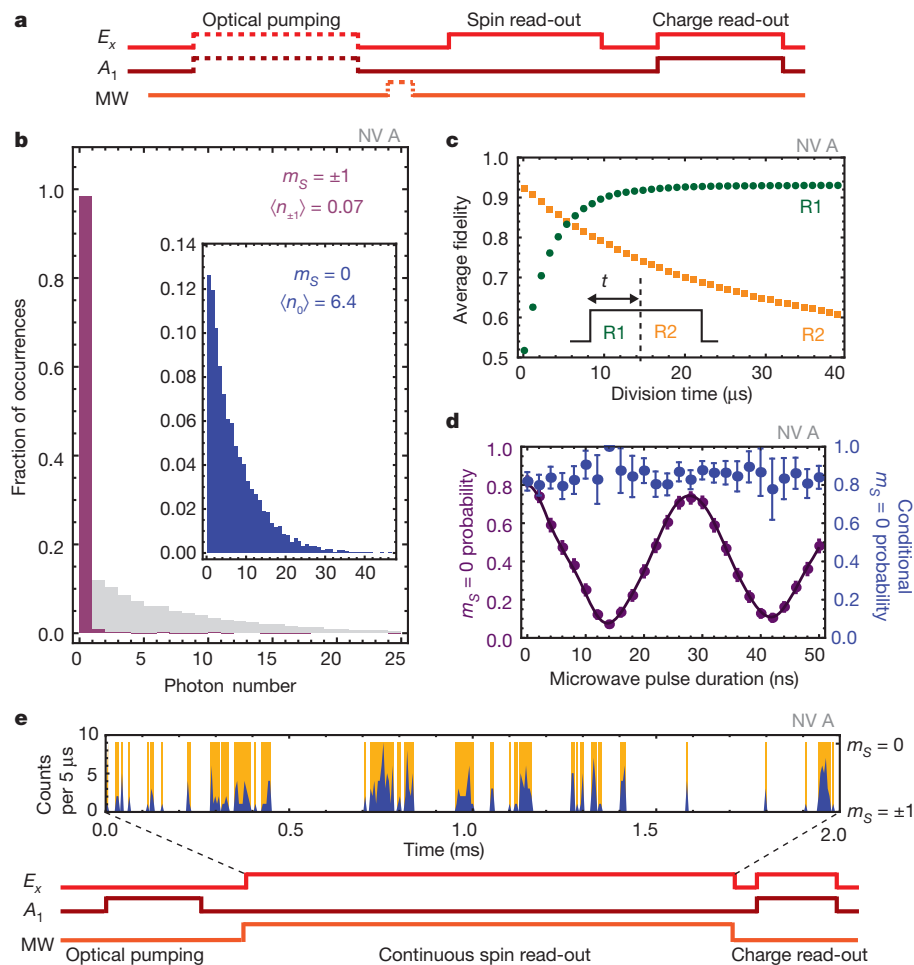


Figure 2 | Projective single-shot read-out of the NV's electronic spin.

a, Pulse sequence used for electronic spin read-out: after charge initialization (532 nm; not shown) the electron is pumped into the $m_S = 0$ (A_1 , dark red) or $m_S = \pm 1$ (E_x , bright red) state, and then undergoes optional microwave spin manipulation and spin read-out by a pulse resonant with E_x . Conditioning on simultaneous resonance during the final charge- and detuning-sensing stage eliminates effects of local electric field fluctuations or ionization (Supplementary Information). **b**, Statistics of photon counts detected during a $t_{ro} = 100$ - μ s electronic spin read-out after initialization into the respective $m_S = \pm 1$ (red) and $m_S = 0$ (superimposed light grey and inset) states, obtained from 10,000 measurement repetitions. **c**, When the 100- μ s read-out pulse is divided into two read-out segments, R1 and R2, with a variable division point, the fidelity of two consecutive segments reaches $83.4 \pm 0.5\%$ for an optimal

division time of 5.5 μ s; the probability of identical sequential outcomes is $82.0 \pm 0.7\%$. Error bars (2 s.e., $n = 10,000$) are smaller than the symbols. **d**, Electronic spin Rabi oscillations between $m_S = 0$ and $m_S = -1$ at axial magnetic field $B_z \approx 15$ G (purple): each data point comes from 1,000 single-shot read-out repetitions. The fit, which includes the detailed hyperfine level structure, yields a visibility of $78 \pm 8\%$, where a maximum of 84% can be expected. Blue data points show the measurement outcome after projection into $m_S = 0$ by selecting only read-out events with photons detected within the first 400 ns (Supplementary Information). All errors and error bars, 2 s.e. **e**, Quantum jumps in the fluorescence time trace during continuous spin read-out. Durations of dark periods depend on the microwave Rabi frequency (Supplementary Information). Blue data indicates the counts per 5- μ s read-out bin, and the deduced spin state is shown in orange.

between measurement outcomes indicate that the read-out is projective. Following preparation of a superposition of spin states, we condition on detection of at least one photon (that is, measurement outcome $m_S = 0$) during a first short read-out pulse, and probe the resulting spin state with a second read-out (Fig. 2d, blue data points). Regardless of the initial spin state, we observe a constant high probability of obtaining $m_S = 0$ in the second read-out. This shows that the read-out method is strongly projective and well suited for application in measurement-based quantum protocols.

We use projective read-out of the electronic spin in combination with quantum gate operations for initialization and read-out of a few-qubit nuclear spin register. We first demonstrate the concept of measurement-based preparation on a single nuclear qubit. The electronic spin resonance spectrum for NV B (Fig. 3a, green trace) reveals the coupling to the host $I = 1$ ^{14}N nuclear spin: two partly overlapping sets of three hyperfine lines correspond to the $m_S = 0 \leftrightarrow -1$ and

$m_S = 0 \leftrightarrow +1$ electronic spin transitions, Zeeman-split by ~ 2 MHz in Earth's magnetic field. The outermost transitions are associated with a specific nuclear spin state with spin projection m_I , for example $(m_S, m_I) = (0, -1) \leftrightarrow (-1, -1)$ at 2.874 GHz. Our initialization procedure works as follows (Fig. 3a, circuit diagram). First we prepare the electronic spin in $m_S = \pm 1$. We then perform a nuclear-spin-controlled NOT operation on the electronic spin by applying a π -pulse at 2.874 GHz; this operation rotates the electronic spin into $m_S = 0$ only when $m_I = -1$. Finally we read out the electronic spin for 400 ns. If one or more photons are detected during this interval, the two-spin system is projected into $(m_S, m_I) = (0, -1)$. Alternatively, if we run the same protocol with initial electronic spin state $m_S = 0$, we prepare the nuclear spin with $m_I = \{0, +1\}$.

The efficiency of the nuclear spin initialization is evidenced by its drastic effect on the electronic spin resonance spectrum (Fig. 3a). Whereas before preparation the depths of the different hyperfine lines

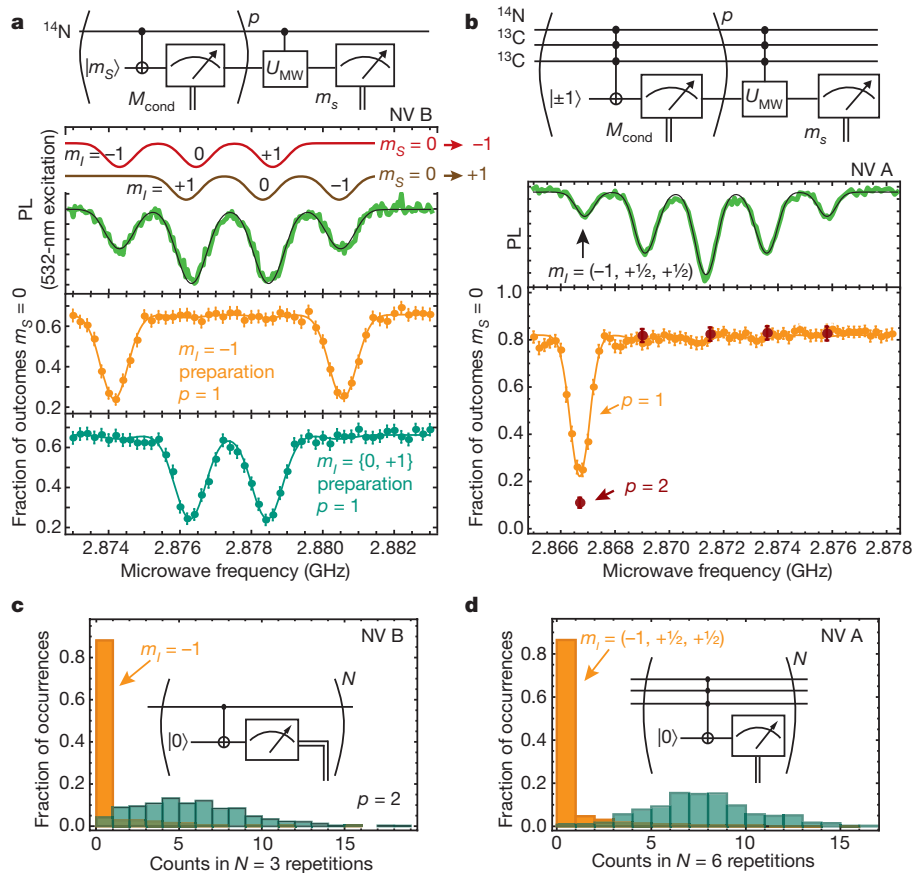


Figure 3 | Nuclear spin preparation and read-out. **a**, Measurement-based preparation of a single ^{14}N nuclear spin. In Earth's ambient magnetic field of ~ 0.5 G, without nuclear spin polarization we observe four resonances in the hyperfine spectrum (green trace) for NV B; the outer two correspond to the nuclear spin state with $m_I = -1$ and the central two are combinations of the states with $m_I = \{0, +1\}$. PL, photoluminescence intensity. Red and brown traces indicate transitions to the $m_S = +1$ and $m_S = -1$ states, respectively. As indicated in the circuit diagram, to initialize the nuclear spin we entangle it with the electronic spin and then read out the latter; p denotes the number of preparation steps, each of which is one iteration of the section of the circuit diagram enclosed in parentheses. M_{cond} indicates the conditioning measurement, M_e the electron spin read-out and U_{MW} the microwave spin manipulation. Data for the $m_I = -1$ preparation is shown in orange; data for the $m_I = \{0, +1\}$ preparation is shown in cyan. Fits to Gaussian spectra show an amplitude ratio of $96 \pm 4\%$ in the desired nuclear spin state. **b**, Measurement-based preparation of a three-nuclear-spin register. Using a similar sequence (circuit diagram), we prepare a well-defined state for all three nuclear spins. A portion of the uninitialized hyperfine spectrum (green) contains 12 partly superposed lines, of which we prepare the single line corresponding to

$m_I = (-1, 1/2, 1/2)$ (orange). Gaussian fits constrained to known hyperfine splittings yield an amplitude ratio of $88 \pm 10\%$. The observed visibility can be improved by performing two preparation steps and electronic spin repumping ($p = 2$; five red data points), yielding a contrast of 82% of the expected visibility from known read-out fidelity (Supplementary Information). Uncertainties and error bars, 2 s.e. ($n = 1,000$ for **a** and **b** with $p = 1$; selected from 10,000 measurement runs for **b** with $p = 2$). **c**, Single-shot measurement of the ^{14}N nuclear spin, preceded by two preparation steps ($p = 2$). Read-out (three repetitions) conditioned on successful preparation distinguishes $m_I = -1$ (orange; threshold < 1 count) from $m_I = \{0, +1\}$ (cyan) with an average fidelity of $92 \pm 2\%$ (Supplementary Information). **d**, Multiple-nuclear-spin read-out. Using a sequence similar to that in **c** (inset), we distinguish one of the 12 hyperfine states associated with NV A. To prepare nuclear spin states, we perform the read-out procedure seven times and keep only data with zero total counts (identified as $m_I = (-1, 1/2, 1/2)$) or ≥ 2 counts per initialization step (other states). Subsequent read-out with six repetitions ($m_I = (-1, 1/2, 1/2)$; state discrimination threshold, < 3 total counts) achieves a $96.7 \pm 0.8\%$ average fidelity for preparation and detection of the nuclear spin configuration.

indicate an equal mixture of the nuclear spin states (green trace), after preparation only the hyperfine lines corresponding to the prepared states are visible ($m_I = -1$ for the orange trace and $m_I = \{0, +1\}$ for the cyan trace).

The same nuclear spin initialization scheme can be applied to multi-qubit registers. Figure 3b shows the electronic spin resonance spectrum of NV A (green trace), whose electronic spin is coupled to both the host ^{14}N nuclear spin and two nearby ^{13}C nuclei (Supplementary Information). The lowest-frequency line corresponds to a single state of the three nuclear spins. A π -pulse on this transition therefore implements a triple-controlled NOT operation on the electronic spin (Fig. 3b, circuit diagram), allowing the initialization of all three nuclear spins (Fig. 3b, orange trace). The initialization can be further improved by repeating the preparation step (Fig. 3b, red data points).

The nuclear qubits can be read out in a single shot by applying a nuclear-spin-controlled NOT operation to the electronic spin and subsequently reading out the electronic spin (Fig. 3c, d, inset circuit diagrams). Because the back-action of the electronic spin measurement on the nuclear spin is weak, we can repeat the process to obtain higher read-out fidelity^{5,8}. Figure 3c compares the resulting photon statistics for NV B after initialization into the single nuclear spin state $m_I = -1$ with those obtained for $m_I = \{0, +1\}$, and indicates an average read-out fidelity of $92 \pm 2\%$. This number is a lower bound on the true read-out fidelity, as it includes errors in state preparation.

A straightforward extension of this scheme can be used to read out the complete state of a register of multiple nuclear spins. By using a

multiply controlled NOT gate in the read-out sequence, we can measure in a single shot whether the register is in a particular configuration. We demonstrate this procedure on NV A, where we identify the three-nuclear-spin state $m_I = (-1, 1/2, 1/2)$ (Fig. 3d and Supplementary Information). The other possible configurations can be probed by sequential application of this read-out scheme to different spectrally resolved hyperfine transitions, or, alternatively, by systematically flipping the nuclear spin qubits and repeating the read-out on the same hyperfine transition.

Electronic–nuclear flip-flop processes in the optically excited state, which reduce the nuclear spin read-out fidelity, pose a significant hurdle to scaling the read-out to more qubits. Critically, resonant read-out allows selection of which states undergo optical excitation. By starting with the electronic spin in an $m_S = \pm 1$ state, optical excitation will occur only when the register is in the state being probed; therefore, no optically induced nuclear spin flips will occur during measurement of any of the other states. Thus, by contrast with schemes depending on off-resonant excitation where each additional read-out step degrades the fidelity, resonant excitation allows scaling of high-fidelity read-out to larger registers.

Finally, we demonstrate the compatibility of all the different techniques discussed here by implementing them in a single experiment: we initialize, coherently manipulate and then read out a two-qubit register consisting of the electronic spin and ^{14}N nuclear spin of NV B. After initializing it to have $(m_S, m_I) = (0, -1)$, we rotate the nuclear spin using a radio-frequency pulse and subsequently rotate the

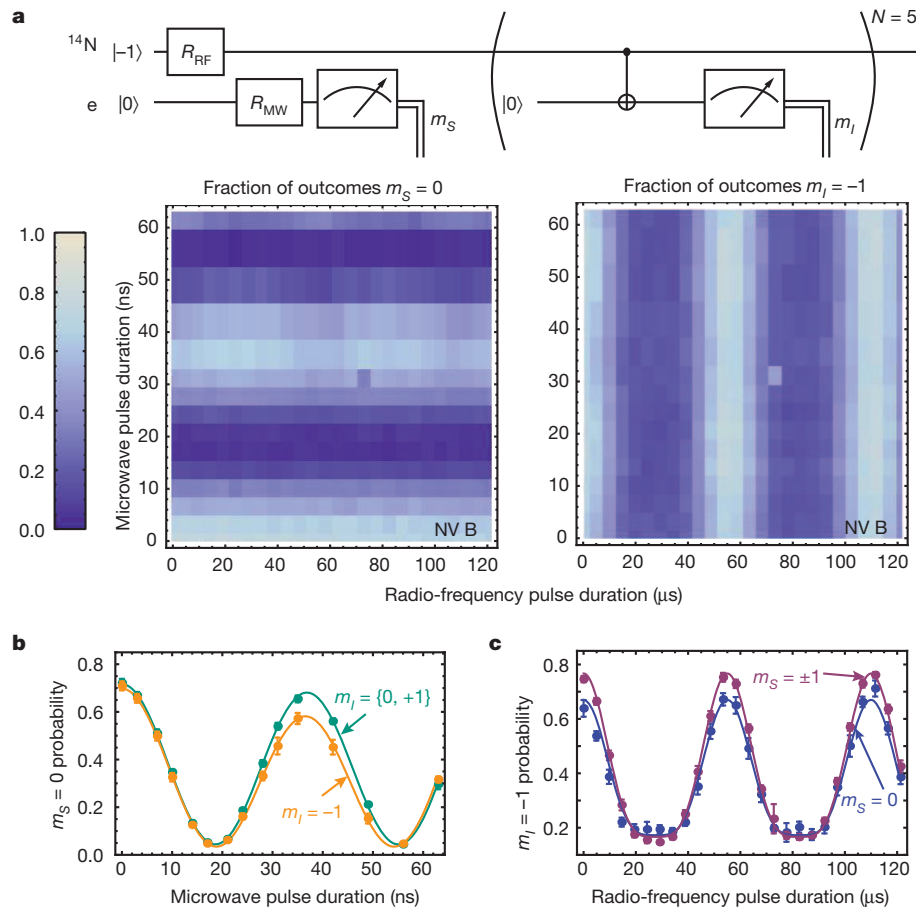


Figure 4 | Initialization, manipulation and read-out of a two-qubit register. **a**, After initialization of NV B into the $(m_S, m_I) = (0, -1)$ state, we use radio-frequency excitation (spin rotation R_{RF} at 4.9464 MHz) to drive the nuclear spin and then microwaves (spin rotation R_{MW} at 2.8774 GHz) to drive the electronic spin. The electronic spin state is subsequently measured for 15 μs , and this is followed by five read-out steps (each of 10 μs) of the ^{14}N nuclear spin state. **b**, Probability of observing $m_S = 0$ conditional on the measured nuclear

spin state and averaged over all radio-frequency pulse durations, as a function of microwave pulse duration (Supplementary Information). **c**, Probability of observing $m_I = -1$ conditional on the observed electronic spin state and averaged over all microwave pulse durations, as a function of radio-frequency pulse duration (Supplementary Information). Error bars and uncertainties, 2 s.e.; data in **a** are based on 1,000 measurements per pixel.

electronic spin with a microwave pulse. We then read out the electronic spin and, subsequently, the ^{14}N nuclear spin state (Fig. 4a, circuit diagram). The left-hand plot in Fig. 4a shows the read-out results for the electronic spin qubit, showing Rabi oscillations as a function of microwave pulse length. By contrast, the read-out results for the nuclear spin qubit (right-hand plot of Fig. 4a) show Rabi oscillations as a function of the radio-frequency pulse length.

To quantify crosstalk, we closely examine correlations between the two measurement outcomes. We observe that the electronic spin Rabi oscillation depends on the outcome of the nuclear read-out (Fig. 4b), but this dependence can be fully accounted for by the finite microwave power used in this experiment (Supplementary Information). The observed correlations thus arise from imperfect manipulation rather than from measurement crosstalk. For the nuclear spin read-out, however, true measurement crosstalk appears: nuclear Rabi oscillation amplitudes decrease when the electronic spin is measured to be in $m_S = 0$ (Fig. 4c) because optical excitation during electronic spin read-out (which only succeeds for $m_S = 0$) induces nuclear spin relaxation (see Supplementary Information for details). This effect can be mitigated by improving the collection efficiency (thus reducing the read-out duration), for example by integrating the NV into an optical cavity. Also, application of moderate magnetic fields can decrease the optically induced nuclear spin relaxation rate by orders of magnitude⁵.

Our results have implications for a broad range of spin-based applications. Single-shot electronic spin read-out can drastically improve NV-based sensors by allowing fast, quantum-projection-limited detection, creating opportunities in low-temperature magnetometry^{31,32}. Extension of nuclear spin preparation techniques to remote nuclei in the spin bath may permit line-narrowing for enhanced sensitivity to d.c. magnetic fields. Furthermore, the preparation, manipulation and single-shot read-out of two spins opens the door to the exploration of two-particle quantum correlations, such as Bell's inequalities, and elementary quantum information processing protocols. Importantly, the techniques we describe are extendable to larger spin registers, and can be combined with precise spin qubit control and dynamical decoupling to give coherence protection^{10–12}. The preparation and read-out fidelities reported here are sufficient for demonstrating measurement-based entanglement generation and quantum teleportation of spin qubits, and for exploring elementary quantum error correction schemes¹. Ultimately, the integration of multi-spin registers with quantum optical channels by means of spin-photon entanglement¹³ may allow their application as few-qubit nodes in long-distance quantum communication protocols or distributed quantum information processing networks.

METHODS

All data were obtained by detecting photons emitted into the phonon sideband (wavelength, 650–750 nm). For photoluminescence excitation spectroscopy, excitation with 5.5-nW red light is applied while microwaves at 2.878 GHz, coupled through an on-chip stripline, drive the electronic spin transitions to prevent optical pumping. Scans are recorded in a single laser frequency sweep at $\sim 200\text{ MHz s}^{-1}$, and are preceded by a 10- μs pulse of 532-nm excitation (50 μW). The green light is necessary to reset the negative-charge state of the NV, which can be photoionized by continuous resonant excitation. For all other experiments, 532-nm-induced spectral diffusion must also be controlled: to ensure that the NV is on resonance with the red excitation lasers, we condition our data on strong fluorescence upon simultaneous E_x and A_1 excitation following the experimental sequence (details in Supplementary Information). All errors and error bars are 2 s.e. statistical uncertainty in the mean (95% confidence interval).

Received 11 June; accepted 28 July 2011.

Published online 21 September 2011.

- Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge Univ. Press, 2000).
- Raussendorf, R., Browne, D. E. & Briegel, H. J. Measurement-based quantum computation on cluster states. *Phys. Rev. A* **68**, 022312 (2003).
- Elzerman, J. M. *et al.* Single-shot read-out of an individual electron spin in a quantum dot. *Nature* **430**, 431–435 (2004).

- Vamivakas, A. N. *et al.* Observation of spin-dependent quantum jumps via quantum dot resonance fluorescence. *Nature* **467**, 297–300 (2010).
- Neumann, P. *et al.* Single-shot readout of a single nuclear spin. *Science* **329**, 542–544 (2010).
- Morello, A. *et al.* Single-shot readout of an electron spin in silicon. *Nature* **467**, 687–691 (2010).
- Giedke, G., Taylor, J. M., D'Alessandro, D., Lukin, M. D. & Imamoglu, A. Quantum measurement of a mesoscopic spin ensemble. *Phys. Rev. A* **74**, 032316 (2006).
- Jiang, L. *et al.* Repetitive readout of a single electronic spin via quantum logic with nuclear spin ancillae. *Science* **326**, 267–272 (2009).
- Balasubramanian, G. *et al.* Ultralong spin coherence time in isotopically engineered diamond. *Nature Mater.* **8**, 383–387 (2009).
- Naydenov, B. *et al.* Dynamical decoupling of a single-electron spin at room temperature. *Phys. Rev. B* **83**, 081201(R) (2011).
- de Lange, G., Wang, Z. H., Ristè, D., Dobrovitski, V. V. & Hanson, R. Universal dynamical decoupling of a single solid-state spin from a spin bath. *Science* **330**, 60–63 (2010).
- Ryan, C. A., Hodges, J. S. & Cory, D. G. Robust decoupling techniques to extend quantum coherence in diamond. *Phys. Rev. Lett.* **105**, 200402 (2010).
- Togan, E. *et al.* Quantum entanglement between an optical photon and a solid-state spin qubit. *Nature* **466**, 730–734 (2010).
- Batalov, A. *et al.* Temporal coherence of photons emitted by single nitrogen-vacancy defect centers in diamond using optical Rabi-oscillations. *Phys. Rev. Lett.* **100**, 077401 (2008).
- Buckley, B. B., Fuchs, G. D., Bassett, L. C. & Awschalom, D. D. Spin-light coherence for single-spin measurement and control in diamond. *Science* **330**, 1212–1215 (2010).
- Robledo, L., Bernien, H., van Weperen, I. & Hanson, R. Control and coherence of the optical transition of single nitrogen vacancy centers in diamond. *Phys. Rev. Lett.* **105**, 177403 (2010).
- Childress, L., Taylor, J. M., Sørensen, A. S. & Lukin, M. D. Fault-tolerant quantum communication based on solid-state photon emitters. *Phys. Rev. Lett.* **96**, 070504 (2006).
- Barrett, S. D. & Kok, P. Efficient high-fidelity quantum computation using matter qubits and linear optics. *Phys. Rev. A* **71**, 060310(R) (2005).
- Jiang, L., Taylor, J. M. & Lukin, M. D. Fast and robust approach to long-distance quantum communication with atomic ensembles. *Phys. Rev. A* **76**, 012301 (2007).
- Dutt, M. V. G. *et al.* Quantum register based on individual electronic and nuclear spin qubits in diamond. *Science* **316**, 1312–1316 (2007).
- Neumann, P. *et al.* Multipartite entanglement among single spins in diamond. *Science* **320**, 1326–1329 (2008).
- Fuchs, G. D. *et al.* Excited-state spectroscopy using single spin manipulation in diamond. *Phys. Rev. Lett.* **101**, 117601 (2008).
- Jacques, V. *et al.* Dynamic polarization of single nuclear spins by optical pumping of nitrogen-vacancy color centers in diamond at room temperature. *Phys. Rev. Lett.* **102**, 057403 (2009).
- Happer, W. Optical pumping. *Rev. Mod. Phys.* **44**, 169–249 (1972).
- Blatt, R. & Zoller, P. Quantum jumps in atomic systems. *Eur. J. Phys.* **9**, 250–256 (1988).
- Tamarat, P. *et al.* Spin-flip and spin-conserving optical transitions of the nitrogen-vacancy centre in diamond. *N. J. Phys.* **10**, 045004 (2008).
- Atature, M. *et al.* Quantum-dot spin-state preparation with near-unity fidelity. *Science* **312**, 551–553 (2006).
- Manson, N. B., Harrison, J. P. & Sellars, M. J. Nitrogen-vacancy center in diamond: model of the electronic structure and associated dynamics. *Phys. Rev. B* **74**, 104303 (2006).
- Fu, K.-M. C. *et al.* Observation of the dynamic Jahn-Teller effect in the excited states of nitrogen-vacancy centers in diamond. *Phys. Rev. Lett.* **103**, 256404 (2009).
- Hadden, J. P. *et al.* Strongly enhanced photon collection from diamond defect centers under microfabricated integrated solid immersion lenses. *Appl. Phys. Lett.* **97**, 241901 (2010).
- Degen, C. L. Scanning magnetic field microscope with a diamond single-spin sensor. *Appl. Phys. Lett.* **92**, 243111 (2008).
- Taylor, J. M. *et al.* High-sensitivity diamond magnetometer with nanoscale resolution. *Nature Phys.* **4**, 810–816 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements L.R. acknowledges support by a Marie Curie Intra European Fellowship within the 7th European Community Framework Programme. L.R., H.B., B.H. and R.H. acknowledge support from the Dutch Organization for Fundamental Research on Matter (FOM) and the European Commission (SOLID). L.C. acknowledges support from Research Corporation for Science Advancement (RCSA).

Author Contributions L.R., L.C. and H.B. conducted the experiments. L.R., L.C., H.B., B.H. and R.H. analysed the data. H.B. and P.F.A.A. fabricated the devices. L.R., L.C. and R.H. wrote the paper. All authors commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.H. (r.hanson@tudelft.nl).

Interannual variability in the oxygen isotopes of atmospheric CO₂ driven by El Niño

Lisa R. Welp¹, Ralph F. Keeling¹, Harro A. J. Meijer², Alane F. Bollenbacher¹, Stephen C. Piper¹, Kei Yoshimura^{1†}, Roger J. Francey³, Colin E. Allison³ & Martin Wahlen¹

The stable isotope ratios of atmospheric CO₂ (¹⁸O/¹⁶O and ¹³C/¹²C) have been monitored since 1977 to improve our understanding of the global carbon cycle, because biosphere–atmosphere exchange fluxes affect the different atomic masses in a measurable way¹. Interpreting the ¹⁸O/¹⁶O variability has proved difficult, however, because oxygen isotopes in CO₂ are influenced by both the carbon cycle and the water cycle². Previous attention focused on the decreasing ¹⁸O/¹⁶O ratio in the 1990s, observed by the global Cooperative Air Sampling Network of the US National Oceanic and Atmospheric Administration Earth System Research Laboratory. This decrease was attributed variously to a number of processes including an increase in Northern Hemisphere soil respiration³; a global increase in C₄ crops at the expense of C₃ forests⁴; and environmental conditions, such as atmospheric turbulence⁵ and solar radiation⁶, that affect CO₂ exchange between leaves and the atmosphere. Here we present 30 years' worth of data on ¹⁸O/¹⁶O in CO₂ from the Scripps Institution of Oceanography global flask network and show that the interannual variability is strongly related to the El Niño/Southern Oscillation. We suggest that the redistribution of moisture and rainfall in the tropics during an El Niño increases the ¹⁸O/¹⁶O ratio of precipitation and plant water, and that this signal is then passed on to atmospheric CO₂ by biosphere–atmosphere gas exchange. We show how the decay time of the El Niño anomaly in this data set can be useful in constraining global gross primary production. Our analysis shows a rapid recovery from El Niño events, implying a shorter cycling time of CO₂ with respect to the terrestrial biosphere and oceans than previously estimated. Our analysis suggests that current estimates of global gross primary production, of 120 petagrams of carbon per year⁷, may be too low, and that a best guess of 150–175 petagrams of carbon per year better reflects the observed rapid cycling of CO₂. Although still tentative, such a revision would present a new benchmark by which to evaluate global biospheric carbon cycling models.

Quantifying global-scale gross primary production (GPP) has been difficult because there are no direct measures at scales greater than the leaf level. Consequently, inferences are made from atmospheric-based CO₂ flux measurements (for example eddy covariance⁷), but these can only measure the net ecosystem exchange, that is, the small residual between GPP and ecosystem respiration. Satellite-based observations estimate GPP on the basis of absorbed photosynthetic radiation⁸. Both approaches rely heavily on biospheric process models to generate GPP estimates. We present a new approach that does not depend on biospheric process models, using existing long-term time series of CO₂ isotopes to quantify mean global GPP over the past several decades.

Stable isotope analysis of CO₂ started with flask samples from the South Pole (SPO) and Christmas Island stations in 1977 and later expanded to include the Alert (ALT), Point Barrow, La Jolla, Mauna Loa (MLO), Cape Kumukahi, American Samoa, Kermadec Islands and Baring Head stations. Figure 1 shows deseasonalized monthly mean

¹⁸O/¹⁶O observations and spline fits at each station expressed using delta notation, $\delta^{18}\text{O} = (^{18}\text{O}/^{16}\text{O})_{\text{sample}} / (^{18}\text{O}/^{16}\text{O})_{\text{VPDB}} - 1$, with the Vienna Pee Dee Belemnite (VPDB) standard. Interannual variability of $\delta^{18}\text{O}$ -CO₂ (that is, $\delta^{18}\text{O}$ of the CO₂ samples) is characterized by oscillations of 2–10 yr, with relatively consistent timing and magnitude of anomalies at all stations and without any persistent long-term trend. The lack of a trend in ¹⁸O/¹⁶O contrasts with the increasing trend in the CO₂ mixing ratio, of roughly 1.5 p.p.m. yr⁻¹, and the decreasing trend in $\delta^{13}\text{C} = (^{13}\text{C}/^{12}\text{C})_{\text{sample}} / (^{13}\text{C}/^{12}\text{C})_{\text{VPDB}} - 1$, of -0.02‰ yr^{-1} , over the same period¹. This is consistent with the fact that fossil fuel emissions contribute little to the ¹⁸O/¹⁶O budget of CO₂, and that oxygen atoms in CO₂ are renewed by exchange with oxygen atoms of water in leaves, soils and sea water approximately every 2 yr (refs 2, 9, 10) or less, as will be discussed below. Fortunately, signal-to-noise ratios are generally much larger for $\delta^{18}\text{O}$ than for $\delta^{13}\text{C}$ in CO₂.

Cross-correlation analysis determined that $\delta^{18}\text{O}$ -CO₂ anomalies at station SPO lag anomalies at the low-latitude station MLO by 4–5 months. Stations ALT and PTB, which are at high northern latitudes, lag MLO by 3–4 months. This suggests that anomalies start in the tropics and propagate to high latitudes over the course of several months. Maximum correlation coefficients between SPO and both MLO and ALT were approximately 0.4 when allowing for lags between stations.

During El Niño/Southern Oscillation (ENSO) events, there are significant changes in the amount and intensity of rainfall in the tropics, especially over Southeast Asia and northern South America¹¹. Changes in the amount of precipitation also affect the $\delta^{18}\text{O}$ value of water through the 'amount effect' and related processes¹². Figure 2 shows the correlation between the ENSO precipitation index¹³ (ESPI) and the $\delta^{18}\text{O}$ value of precipitation as hindcast by the IsoGSM model¹⁴ driven by reanalysis winds. During El Niño years, the $\delta^{18}\text{O}$ value of precipitation (δ_{ppt}) increases in regions where the amount of precipitation decreases, such as Southeast Asia and northern South America. These are also regions with very high annual net primary production (NPP) and GPP. Figure 3a, b shows that $\delta^{18}\text{O}$ -CO₂ is strongly positively correlated with δ_{ppt} in the tropics, whereas Fig. 3c, d shows that in the same regions it is negatively correlated with relative humidity. This further supports a link between $\delta^{18}\text{O}$ -CO₂ and ENSO variability through the tropical water cycle.

The interannual variability of $\delta^{18}\text{O}$ -CO₂ measured by the Scripps Institution of Oceanography (SIO) flask network was confirmed by comparison with an independent record of similar length from Cape Grim (CGO) station as measured by the Commonwealth Scientific and Industrial Research Organisation¹⁵ (CSIRO), also shown in Fig. 1. CGO and SPO have very similar correlations with ENSO forcing, lagging ENSO by 8–11 months with maximum correlation coefficients of 0.4. SIO flask observations were also compared with those of the US National Oceanic and Atmospheric Administration Earth System Research Laboratory (Supplementary Information, section 1).

¹Scripps Institution of Oceanography, University of California San Diego, 9500 Gilman Drive, La Jolla, California 92093-0244, USA. ²Center for Isotope Research, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands. ³CSIRO Marine and Atmospheric Research, PB 1, Aspendale, Victoria 3195, Australia. [†]Present address: Atmosphere and Ocean Research Institute, University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8568, Japan.

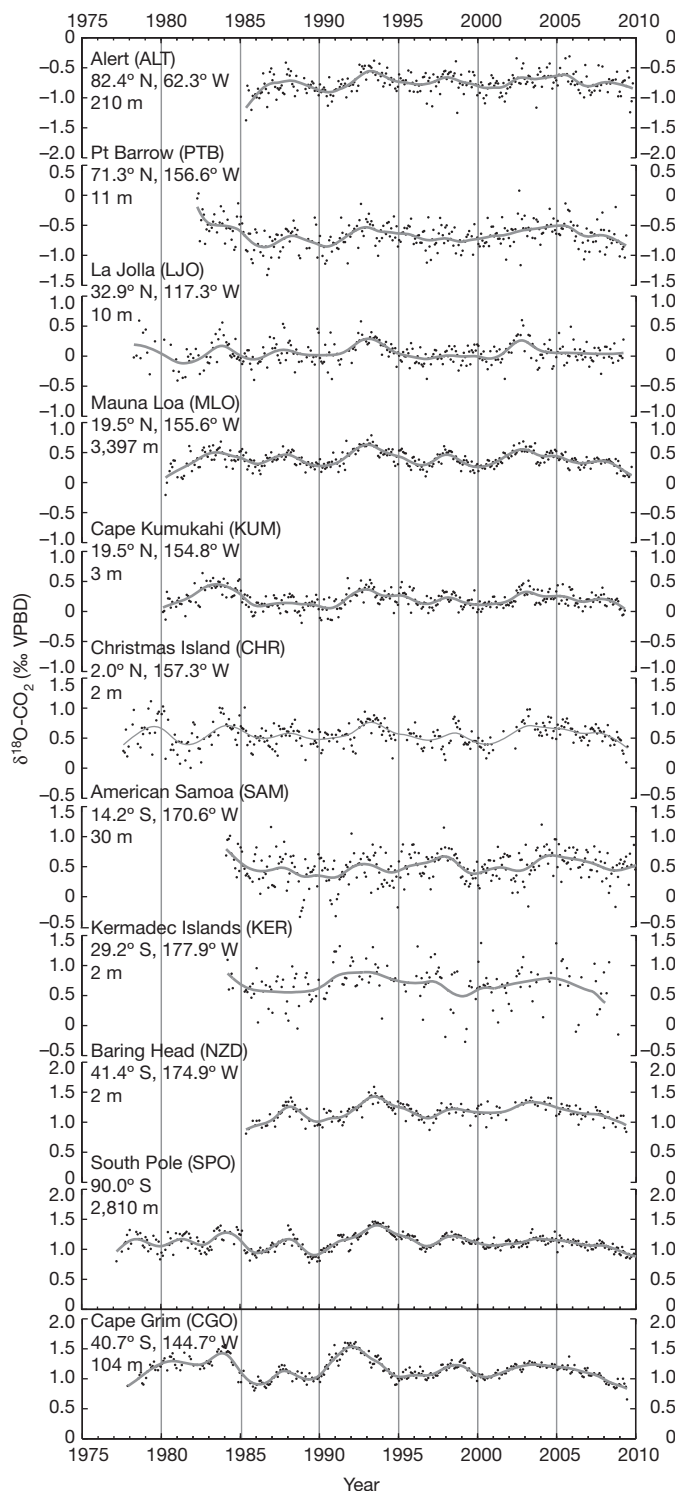


Figure 1 | Measurements of $\delta^{18}\text{O}-\text{CO}_2$ from the SIO flask network and CSIRO. Monthly mean deseasonalized $\delta^{18}\text{O}-\text{CO}_2$ station data (black dots) with long-term spline fits (grey lines). All stations (latitude, longitude and altitude as shown) are from the SIO flask network with the exception of the CSIRO station CGO, at the bottom of the figure. A seasonal harmonic fit was subtracted from the monthly means to produce the deseasonalized observations presented here. The 1σ mass spectrometer precision of both laboratories since 1990 is $\sim 0.014\text{‰}$. The SIO estimates that the 1σ error of duplicate flask measurements is $\sim 0.025\text{‰}$. Measurement uncertainties for both laboratories are larger before 1990.

We used a simple empirical model to demonstrate that the inter-annual variability in $\delta^{18}\text{O}-\text{CO}_2$ is closely related to a low-pass filter of ENSO. The model follows from mass balance considerations (see

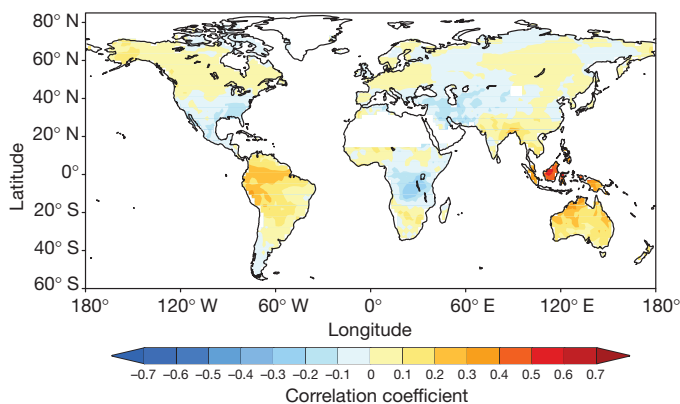


Figure 2 | Correlations between precipitation $\delta^{18}\text{O}$ from IsoGSM and ENSO. We use the ENSO precipitation index (ESPI) as a proxy for ENSO variability. Positive correlation indicates that precipitation is enriched in ^{18}O during the El Niño phase.

Supplementary Information, section 2, for a full derivation), treating the troposphere as two well-mixed boxes, one for the Northern Hemisphere and one for the Southern Hemisphere (N and S, respectively), assuming a linear ENSO forcing on the $\delta^{18}\text{O}-\text{CO}_2$ values for each hemisphere (δ'_N and δ'_S , respectively) through the influence of the terrestrial biosphere and incorporating non-ENSO background processes that effectively damp the ENSO isotope anomalies:

$$\frac{d\delta'_N}{dt} = f_N \times A \times \text{ENSO}'(t - \text{lag}) - \frac{\delta'_N}{\tau_N} + \frac{\delta'_S - \delta'_N}{\tau_{\text{mix}}} \quad (1)$$

$$\frac{d\delta'_S}{dt} = (1 - f_N) \times A \times \text{ENSO}'(t - \text{lag}) - \frac{\delta'_S}{\tau_S} + \frac{\delta'_N - \delta'_S}{\tau_{\text{mix}}} \quad (2)$$

Fitting parameters include f_N , A , τ_N , τ_S and lag, where f_N represents the fraction of the total ENSO forcing within the Northern Hemisphere; A is the scaling term in units of $\text{‰ ESPI}^{-1} \text{yr}^{-1}$, and represents the conversion of the ESPI anomalies from the mean (ENSO') into $\delta^{18}\text{O}-\text{CO}_2$ anomalies from the mean (δ'_N or δ'_S , depending on hemisphere); τ_N and τ_S are the Northern Hemisphere and Southern

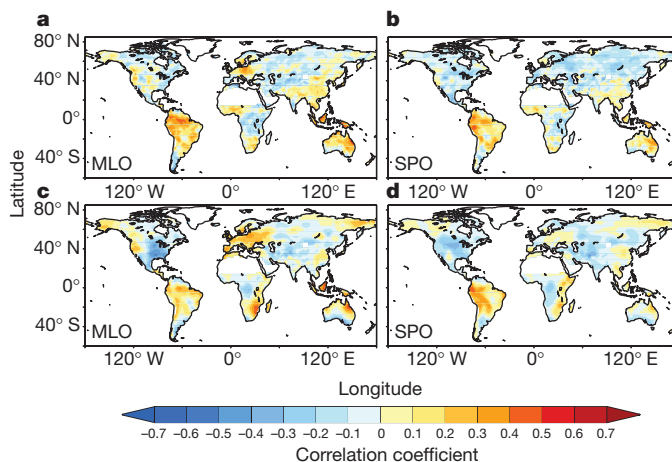


Figure 3 | Correlations between flask $\delta^{18}\text{O}-\text{CO}_2$ records and precipitation $\delta^{18}\text{O}$ and relative humidity from IsoGSM. a, b, Correlation between the previous 9-month running mean of precipitation $\delta^{18}\text{O}$ and the monthly mean deseasonalized $\delta^{18}\text{O}-\text{CO}_2$ from station MLO (a) and station SPO (b). Positive correlations indicate that $\delta^{18}\text{O}-\text{CO}_2$ increases when precipitation $\delta^{18}\text{O}$ increases. c, d, Correlation between the NPP-weighted, 9-month running mean of negative relative humidity and deseasonalized $\delta^{18}\text{O}-\text{CO}_2$ from MLO (c) and SPO (d). Positive correlations indicate that $\delta^{18}\text{O}-\text{CO}_2$ increases when relative humidity decreases. NPP-weighted relative humidity was used here to give weight to months with higher vegetation productivity when leaf-water isotope anomalies are passed onto CO_2 .

Hemisphere turnover times of oxygen atoms in CO_2 ; and lag is the temporal offset between ESPI and the isotopic ENSO forcing. The constant τ_{mix} , which represents the interhemispheric mixing time, was set to a value of 12 months.

The box model approach assumes that surface $\delta^{18}\text{O}\text{-CO}_2$ measurements represent the entire tropospheric column. Although aircraft measurements have observed variability within the troposphere, the fact that the seasonal cycle of $\delta^{18}\text{O}\text{-CO}_2$ observed at MLO is similar to that in the upper troposphere¹⁶ means that the box model may be a sufficient first attempt because we are tracking a signal that originates in the tropics and is transported vertically before reaching the mid latitudes.

We alternately used MLO and ALT to represent the Northern Hemisphere box, and SPO and CGO for the Southern Hemisphere box (Supplementary Information, section 3). We repeated the fits using 1,000 hypothetical time series with the same autocorrelation as the ENSO index but with random phase¹⁷. All station combinations gave fits to the true ENSO index that were better than a large majority of fits using the hypothetical indices, more than 98% of cases using the SIO station combinations and more than 93% using SIO–CSIRO station combinations (Supplementary Information, section 4). The MLO–SPO combination of data from 1980 to 2009 yielded the best fit between the ENSO model and observations (Fig. 4).

The model fit yielded a short turnover time of $\tau_N = 0.4\text{--}0.8$ yr for $\delta^{18}\text{O}\text{-CO}_2$ in the Northern Hemisphere, depending on the stations used in the analysis and with 1σ errors not greater than 0.3 yr. Sensitivity analysis of the model parameter fit covariance suggested that τ_S is greater than 2 yr, but the upper bound was difficult to determine because of the 1-yr interhemispheric mixing time (τ_{mix}). We expected τ_N to be less than τ_S given the larger land biosphere and the greater exchange of CO_2 with leaf and soil water in the north. Estimated turnover times for the two hemispheres based on gross exchanges with the ocean¹⁸ and the terrestrial biosphere¹⁹ indicated that the Southern Hemisphere turnover time is $\sim 120\%$ longer than that of the Northern Hemisphere. The model-derived, hemisphere-specific turnover times correspond to a global tropospheric mean of

0.7–1.4 yr. Allowing for the 22% of the atmospheric mass contained in the stratosphere²⁰, which was not included in the model, the estimated global atmospheric turnover time increases to 0.9–1.7 yr.

These atmospheric observations suggest that the turnover time for oxygen atoms in CO_2 may be slightly lower than previously published estimates. Turnover times of oxygen in CO_2 of 1.7 yr (ref. 2) and 1.5 yr (ref. 10) have been estimated from bottom-up flux estimates using process models. These estimates are sensitive to assumed values for GPP and the ratio of CO_2 concentration inside the leaf relative to the atmosphere. Our turnover time estimate using $\delta^{18}\text{O}\text{-CO}_2$ anomalies is independent of these assumptions. Recent efforts to balance the carbonyl sulphide budget also point to more exchange of CO_2 between the atmosphere and land vegetation than previously estimated²¹. Field studies have indicated that carbonic anhydrase in soils catalyses the direct exchange of oxygen isotopes between soil water and CO_2 . A recent global estimate of this soil equilibration flux could bring the global turnover time of ref. 10 down as low as ~ 0.9 yr (ref. 22).

The parameter A provides insight into the mechanisms linking $\delta^{18}\text{O}\text{-CO}_2$ to ENSO climate anomalies. The fits yielded A values of 0.27 to $0.35\text{‰ ESPI}^{-1}\text{yr}^{-1}$, depending on the stations used, with 1σ errors of less than $0.08\text{‰ ESPI}^{-1}\text{yr}^{-1}$. To assess which processes contribute most to A , we carried out a scale analysis of the mass balance of ^{18}O in CO_2 (Supplementary Information, section 5). We identified two dominant processes. The first was a change in the δ_{ppt} value associated with ENSO events. Precipitation $\delta^{18}\text{O}$ anomalies are directly transmitted to soil and leaf water, and hence contribute to the $\delta^{18}\text{O}$ value of CO_2 . Generally, the δ_{ppt} value on land in the tropics increases during El Niño events, as evidenced by isotopic studies of precipitation^{23,24}, tropical ice cores²⁵, tree rings²⁶ and isotope-enabled general circulation models¹⁴. The second dominant effect was an increase in the isotopic enrichment of leaf water relative to soil water caused by lower relative humidity during warm and dry El Niño events in the tropics than in La Niña years²⁷.

We estimate that the effects of δ_{ppt} and relative humidity are similar in magnitude and reinforce each other. The magnitude of A can be accounted for if we assume that anomalies in δ_{ppt} and relative humidity influenced 40% of the global land CO_2 fluxes. NPP from 20°S to 20°N is about 40% of global NPP¹⁹, lending support that the global anomalies of $\delta^{18}\text{O}$ in CO_2 may be generated by the tropical land biosphere that is modulated by ENSO. Our analysis showed that many other possible processes were less important, including changes in CO_2 fluxes (for example those related to drought) and fire emissions. Our analysis neglected ENSO-related changes in air–sea CO_2 exchange, in stratosphere–troposphere exchange and in diffuse light, which are also likely to be small.

The model-derived fraction of isotopic ENSO forcing injected into the Northern Hemisphere (f_N) was 0.49–0.76 depending on the station combinations used. Lags between ESPI and atmospheric $\delta^{18}\text{O}\text{-CO}_2$ forcing were small, ranging from -2 months to 1 month, and had 1σ errors of 1 month. The model thus successfully accounts for the observed lag between ESPI and $\delta^{18}\text{O}\text{-CO}_2$ through the finite hemispheric turnover times.

We can relate our model-derived turnover time to the gross flux of CO_2 equilibrating with the ocean and land surface water. This gross exchange flux can be used to estimate GPP by subtracting air–sea exchange; the amount of CO_2 that escapes leaf interiors without being fixed into organic matter, which is directly related to the ratio of CO_2 inside the leaf to that in the atmosphere; and the soil invasion flux (Supplementary Information, section 6). Reconciling the short turnover time (0.9–1.7 yr) with current understanding of the global carbon budget would require that the soil invasion CO_2 flux be near the upper case considered in ref. 22 and much greater than soil respiration. At present, there are not enough field observations to confirm such high fluxes of soil invasion. Alternatively, and more plausibly, the fast response can be accounted for by revising global GPP upwards from 120 Pg C yr^{-1} (ref. 7) to $150\text{--}175\text{ Pg C yr}^{-1}$. Using this approach, we

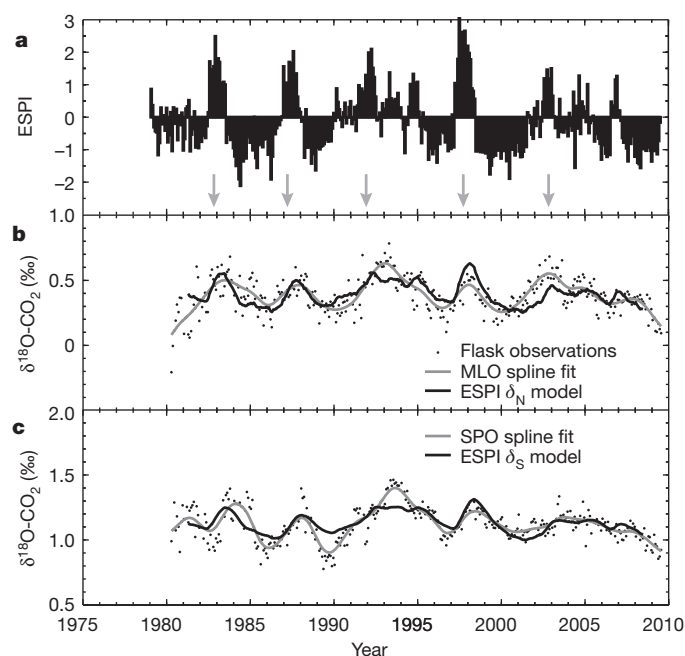


Figure 4 | ENSO index and two-box model results. **a**, ESPI was used as a proxy for ENSO variability. Arrows denote El Niño events. **b**, Empirical ENSO model fit for the Northern Hemisphere (δ_N , black line), compared with the deseasonalized monthly flask observations (black dots) and spline fit or interannual variability at station MLO (grey line). **c**, Same as in **b** but for the Southern Hemisphere model fit (δ_S) and station SPO.

argue that 120 Pg C yr^{-1} may actually be a lower bound on GPP. (See Supplementary Information, section 6, for more discussion.)

The timescale of the damping of the ENSO $\delta^{18}\text{O}$ - CO_2 signal provides a constraint on global GPP without the need for detailed isotopic modelling of multiple surface water pools. We have achieved with $\delta^{18}\text{O}$ - CO_2 what was believed to be possible only by combining ^{17}O and ^{18}O data from CO_2 (ref. 28). These results can be used to validate biospheric components of carbon cycle models used to predict future climate change. The measurements and modelling presented here highlight the significance of $\delta^{18}\text{O}$ - CO_2 long-term records as integrators of the carbon and water cycles.

METHODS SUMMARY

Air samples were collected, without drying, in evacuated 5-l glass flasks with a single glass stopcock lubricated with Apiezon grease, and were returned to the laboratory for analysis. Until 1992, isotopic analysis was performed at the CIO University of Groningen, The Netherlands; since then, it has been done at the SIO, La Jolla, California. Offsets between the laboratories have been investigated and corrected²⁹. We curve fitted the daily mean of replicate flask samples with spline and seasonal harmonic components, and monthly means were constructed from observations adjusted to the fifteenth day of each month³⁰. The interannual variability is simply the spline component of the curve fit. Deseasonalized monthly means were calculated by subtracting the seasonal harmonic component from the monthly mean observations.

Parameter fitting of the two-box model was done using a nonlinear least-squares data fitting function in MATLAB ('lsqnonlin') to find the minimum of the sum of squares of the differences between deseasonalized monthly mean flask observations from one station in each hemisphere and the model predictions from equations (1) and (2), weighted by $1/\sigma$ for each station. Estimates of the 1σ error for each station's measurements were not obtained from the mass spectrometer precision but rather from the standard deviation of the residuals between the deseasonalized monthly mean flask observations and the spline fits, which were $\sim 0.08\%$.

Isotope data of CO_2 from the SIO network is publicly available at http://scrippsco2.ucsd.edu/data/atmospheric_co2.html.

Received 22 July 2010; accepted 4 August 2011.

- Keeling, C. D. *et al.* in *A History of Atmospheric CO_2 and Its Effects on Plants, Animals, and Ecosystems* (eds Ehleringer, J. R., Cerling, T. E. & Dearing, M. D.) 83–113 (Springer, 2005).
- Farquhar, G. D. *et al.* Vegetation effects on the isotope composition of oxygen in atmospheric CO_2 . *Nature* **363**, 439–443 (1993).
- Ishizawa, M., Nakazawa, T. & Higuchi, K. A multi-box model study of the role of the biospheric metabolism in the recent decline of $\delta^{18}\text{O}$ in atmospheric CO_2 . *Tellus B* **54**, 307–324 (2002).
- Gillon, J. & Yakir, D. Influence of carbonic anhydrase activity in terrestrial vegetation on the ^{18}O content of atmospheric CO_2 . *Science* **291**, 2584–2587 (2001).
- Lee, X. H. *et al.* Canopy-scale kinetic fractionation of atmospheric carbon dioxide and water vapor isotopes. *Glob. Biogeochem. Cycles* **23**, GB1002 (2009).
- Still, C. J. *et al.* Influence of clouds and diffuse radiation on ecosystem-atmosphere CO_2 and CO^{18}O exchanges. *J. Geophys. Res.* **114**, G01018 (2009).
- Beer, C. *et al.* Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate. *Science* **329**, 834–838 (2010).
- Zhao, M. S., Heinsch, F. A., Nemani, R. R. & Running, S. W. Improvements of the MODIS terrestrial gross and net primary production global data set. *Remote Sens. Environ.* **95**, 164–176 (2005).
- Ciais, P. *et al.* A three-dimensional synthesis study of $\delta^{18}\text{O}$ in atmospheric CO_2 . 1. Surface fluxes. *J. Geophys. Res.* **102**, 5857–5872 (1997).
- Cuntz, M., Ciais, P., Hoffmann, G. & Knorr, W. A comprehensive global three-dimensional model of $\delta^{18}\text{O}$ in atmospheric CO_2 : 1. Validation of surface processes. *J. Geophys. Res.* **108**, 4527 (2003).
- Dai, A. & Wigley, T. M. L. Global patterns of ENSO-induced precipitation. *Geophys. Res. Lett.* **27**, 1283–1286 (2000).
- Gat, J. R., Mook, W. G. & Meijer, H. A. J. *Environmental Isotopes in the Hydrological Cycle, Volume II: Atmospheric Water* 55, 56 (UNESCO/IAEA, 2001).
- Curtis, S. & Adler, R. ENSO indices based on patterns of satellite-derived precipitation. *J. Clim.* **13**, 2786–2793 (2000).
- Yoshimura, K., Kanamitsu, M., Noone, D. & Oki, T. Historical isotope simulation using reanalysis atmospheric data. *J. Geophys. Res.* **113**, D19108 (2008).
- Allison, C. E. & Francey, R. J. Verifying Southern Hemisphere trends in atmospheric carbon dioxide stable isotopes. *J. Geophys. Res.* **112**, D21304 (2007).
- Assonov, S. S., Brenninkmeijer, C. A. M., Schuck, T. J. & Taylor, P. Analysis of ^{13}C and ^{18}O isotope data of CO_2 in CARIBIC aircraft samples as tracers of upper troposphere/lower stratosphere mixing and the global carbon cycle. *Atmos. Chem. Phys.* **10**, 8575–8599 (2010).
- Ebisuzaki, W. A method to estimate the statistical significance of a correlation when the data are serially correlated. *J. Clim.* **10**, 2147–2153 (1997).
- Takahashi, T. S. *et al.* Climatological mean and decadal changes in surface ocean pCO_2 , and net sea-air CO_2 flux over the global oceans. *Deep-Sea Res. II* **56**, 554–577 (2009).
- Randerson, J. T., Thompson, M. V., Conway, T. J., Fung, I. Y. & Field, C. B. The contribution of terrestrial sources and sinks to trends in the seasonal cycle of atmospheric carbon dioxide. *Glob. Biogeochem. Cycles* **11**, 535–560 (1997).
- Appenzeller, C., Holton, J. R. & Rosenlof, K. H. Seasonal variation of mass transport across the tropopause. *J. Geophys. Res.* **101**, 15071–15078 (1996).
- Montzka, S. A. *et al.* On the global distribution, seasonality, and budget of atmospheric carbonyl sulfide (COS) and some similarities to CO_2 . *J. Geophys. Res.* **112**, D09302 (2007).
- Wingate, L. *et al.* The impact of soil microorganisms on the global budget of $\delta^{18}\text{O}$ in atmospheric CO_2 . *Proc. Natl Acad. Sci. USA* **106**, 22411–22415 (2009).
- Cobb, K. M., Adkins, J. F., Partin, J. W. & Clark, B. Regional-scale climate influences on temporal variations of rainwater and cave dripwater oxygen isotopes in northern Borneo. *Earth Planet. Sci. Lett.* **263**, 207–220 (2007).
- Vuille, M. & Werner, M. Stable isotopes in precipitation recording South American summer monsoon and ENSO variability: observations and model results. *Clim. Dyn.* **25**, 401–413 (2005).
- Thompson, L. G. Ice core evidence for climate change in the Tropics: implications for our future. *Quat. Sci. Rev.* **19**, 19–35 (2000).
- Evans, M. N. Toward forward modeling for paleoclimatic proxy signal calibration: a case study with oxygen isotopic composition of tropical woods. *Geochim. Geophys. Geosyst.* **8**, Q07008 (2007).
- Willett, K. M., Jones, P. D., Gillett, N. P. & Thorne, P. W. Recent changes in surface humidity: development of the HadCRUH dataset. *J. Clim.* **21**, 5364–5383 (2008).
- Hoag, K. J., Still, C. J., Fung, I. Y. & Boering, K. A. Triple oxygen isotope composition of tropospheric carbon dioxide as a tracer of terrestrial gross carbon fluxes. *Geophys. Res. Lett.* **32**, L02802 (2005).
- Bollenbacher, A. F. *et al.* Calibration Methodology for the Scripps $^{13}\text{C}/^{12}\text{C}$ and $^{18}\text{O}/^{16}\text{O}$ Stable Isotope Program 1992–1996: a Report Prepared for the Global Environmental Monitoring Program of the World Meteorological Organization (Scripps Institution of Oceanography, 2000).
- Keeling, C. D. *et al.* Exchanges of Atmospheric CO_2 and $^{13}\text{CO}_2$ with the Terrestrial Biosphere and Oceans from 1978 to 2000. I. Global Aspects. SIO Reference No. 01–06 (Scripps Institution of Oceanography, 2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors thank S. Walker for programming assistance and all those involved with flask collection and analysis. This work was supported by the US National Science Foundation (NSF) under grant ATM06-32770, by the US Department of Energy (DOE) under grant DE-SC0005099 and by the US National Aeronautics and Space Administration (NASA) under grant NNX11AF36G. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF, DOE or NASA.

Author Contributions L.R.W. analysed the data. R.F.K. supervised the project. L.R.W. and R.F.K. wrote the paper. H.A.J.M., A.F.B., R.J.F., C.E.A. and M.W. provided data. K.Y. provided the IsoGSM output. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to L.R.W. (lwelp@ucsd.edu) or R.F.K. (rkeeling@ucsd.edu).

Transcriptomic analysis of avian digits reveals conserved and derived digit identities in birds

Zhe Wang¹, Rebecca L. Young¹, Huiling Xue² & Günter P. Wagner¹

Morphological characters are the result of developmental gene expression. The identity of a character is ultimately grounded in the gene regulatory network directing development and thus whole-genome gene expression data can provide evidence about character identity. This approach has been successfully used to assess cell-type identity^{1–3}. Here we use transcriptomic data to address a long-standing uncertainty in evolutionary biology, the identity of avian wing digits^{4,5}. Embryological evidence clearly identifies the three wing digits as developing from digit positions 2, 3 and 4 (ref. 6), whereas palaeontological data suggest that they are digits I, II and III⁷. We compare the transcriptomes of the wing and foot digits and find a strong signal that unites the first wing digit with the first foot digit, even though the first wing digit develops from embryological position 2. Interestingly, our transcriptomic data of the posterior digits show a higher degree of differentiation among forelimb digits compared with hindlimb digits. These data show that in the stem lineage of birds the first digit underwent a translocation from digit position 1 to position 2, and further indicate that the posterior wing digits have unique identities contrary to any model of avian digit identity proposed so far^{5,8}.

To address the problem of avian digit identity we performed mRNA-seq analysis from pools of 40 individually isolated chicken forelimb and hindlimb digits at two stages of development: stages 28/29 and 31 (Fig. 1a–d). A total of 14,692 annotated genes was detected and compared among samples. Sampling pools of populations are highly reproducible (Supplementary Fig. 1). In describing the results we distinguish nominal digit identities, based on position, from biological digit identities; that is, those that indicate homologies. Nominal digit identities use a three-letter code, the first giving developmental stage (E, early for stage 28/29; L, late for stage 31) and the second giving limb identity (F, forelimb; H, hindlimb). The two letters are followed by a lowercase letter giving the anterior–posterior location of chondrified digits, with ‘a’ for the most anterior digit and ‘b’, ‘c’ and ‘d’ for the posterior digits. In contrast, biological digit identities (homologies) are indicated by roman numerals I, II, III and IV. Finally, embryonic digit positions are indicated with Arabic numerals from anterior to posterior: 1, 2, and so on. These terminological differences are necessary to distinguish descriptive references to certain digits in the adult hand (a, b, etc.) from statements about embryological digit positions (1, 2, etc.) and biological digit identity or homology (digits I, II, etc.).

Transcriptomic data have not been systematically applied to address questions of digit homology. At the most general level, the transcriptome is a phenotypic character and is thus a legitimate source of evidence regarding character identity. Gene expression data from a limited number of genes have successfully been applied to questions of homology in the past^{9–11}. Cell-type identity is another area where gene expression has proven to be valuable^{12–14}. Here we refer to homology as the historical continuity of the developmental programme that determines developmental individuality of the character. By this definition,

homology is thus tied to continuity in the expression of regulatory genes^{15,16}.

First we performed a multidimensional scaling (MDS) analysis¹⁷ to identify the overall similarity structure of the transcriptomes (Fig. 1e). After removing the effect of forelimb–hindlimb differences—that is, removing differentially expressed genes between forelimb and hindlimb—the ordination analysis reveals four clearly resolved clusters corresponding to early and late anterior digit clusters, and two clusters of early and late posterior digits. The transcriptomic data thus contain two strong signals: one pertaining to the difference between the first digits and the posterior digits, and the other reflecting developmental stages.

The heat map of correlation coefficients between samples and hierarchical clustering analyses based on the expression of all detected genes also give strong support for a cluster of anterior digits (Fig. 1f and Supplementary Fig. 2). These data indicate that the anterior digits in the wing and the hindlimb are homologous (Fa = Ha). There is broad agreement that the anterior-most digit in the chicken foot is homologous to digit I in the ancestral five-digit limb of crown-group tetrapods (Ha = DI). We conclude that the most anterior digit in the wing is likely to be homologous to digit I. The transcriptomic data are consistent with the morphological data that show a continuity of digit I identity along the theropod lineage⁷.

The existence of a cluster that unites the forelimb and hindlimb digits I is significant because the first forelimb digit in the chick develops from the embryonic position that normally develops into digit II, that is, Fa develops from embryonic position F2 whereas Ha develops from embryonic position H1 (ref. 18). The difference between the digit I cluster and the posterior digits could reflect the posterior to anterior gradient in digit development¹⁹. However, the separation of the anterior digit I cluster is also found in stage 31, where the anterior and the posterior digits are at comparable levels of differentiation (compare Fig. 1c, d). To analyse which genes are driving the digit I cluster, we identified 556 significantly differentially expressed genes between digit I and the other digits (Supplementary Fig. 3). Gene Ontology (GO) analysis of these differentially expressed genes detected seven enriched GO terms of molecular function and biological process. Transcription factor activity is the top GO term, indicating that the transcriptome data sort digits according to their gene regulatory signature (Supplementary Table 1). Other over-represented GO terms are DNA binding, regulation of transcription, and skeletal system development. These results indicate that the distinction between the first digits and the rest of the digits reflects the developmental individuality of digit I relative to the posterior digits; that is, Fa and Ha are biologically equivalent or true homologues, as suggested by palaeontological⁷ and *in situ* hybridization (ISH) studies^{20–22}. This result suggests that, in the theropod hand, the embryological position of digit I has changed from position 1, as in the ancestor of amniotes, to position 2 in the stem lineage of birds, consistent with the frame shift hypothesis of avian digit identity⁷.

To gain more insight into the genetic nature of digit I identity we searched for digit I markers and identified *Hoxd12*, *Hand2*, *Zic3* and

¹Yale Systems Biology Institute, and Department of Ecology and Evolutionary Biology, Yale University, West Haven, Connecticut 06516, USA. ²Department of Genetics, Yale University School of Medicine, New Haven, Connecticut 06510, USA.

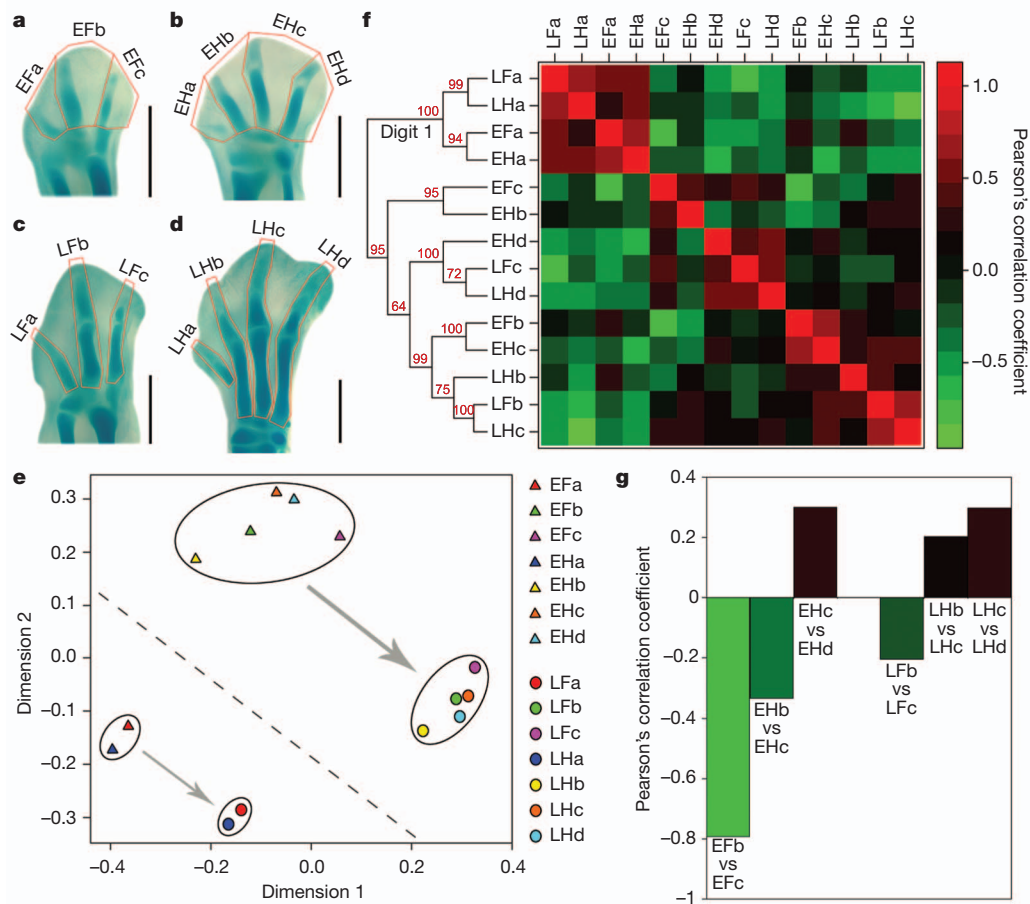


Figure 1 | Chicken embryonic digit transcriptomes identify forelimb and hindlimb digits I but no clear correspondence between posterior forelimbs and hindlimbs. a–d, Alcian-blue-stained autopods. Red lines indicate the dissection protocol used to obtain the samples for the RNA-seq analysis. Scale bar, 1 mm. a, b, Stage 28; c, d, stage 31. e, Ordination of transcriptome

Lhx9, *Hoxd12* and *Hand2* are well known negative markers of the anterior-most digit^{20,23,24} (Supplementary Figs 4 and 5 and Supplementary Discussion). We found two novel positive markers of digit I: *Zic3* and *Lhx9* (Figs 2 and Supplementary Fig. 6).

Zic3 is always expressed more strongly in the first digit than the posterior digits, in particular at stages 26 and 28 (Fig. 2). *Zic3* is a transcriptional co-factor that physically interacts with Gli3, which is important in the transduction of the Shh signal²⁵. Digit I has been proposed to be refractory to Shh signalling¹⁹, and the interaction of *Zic3* protein with Gli3 may have an involvement in this (Supplementary Discussion). *Lhx9* is a LIM homeodomain transcription factor that controls three-dimensional limb patterning in mouse²⁶ (Supplementary Discussion).

The similarity analysis of digit transcriptomes presented above does not determine to which hindlimb digits the second and third wing digits correspond. The inspection of the cluster analysis and the correlation coefficients show a pattern of conflicting signals (Fig. 1f and Supplementary Fig. 7). There are two well supported clusters that unite the second wing digit with the third hindlimb digit, (EFb, EHc) and (LFb, LHc), but these two clusters are united with the late second hindlimb digit, LHb. There is a cluster (EHd, LHd, LFc) that is indicative of the third digit in the wing having an affinity to the fourth hindlimb, but there is also a cluster that unites the early third forelimb digit with the early second hindlimb digit (EFc, EHb), indicating an incoherent signal. To identify the source of this ambiguity we propose that the degree of differentiation among the posterior digits might differ between forelimbs and hindlimbs.

similarity, after removing forelimb–hindlimb signals. f, Heat map of Pearson's correlation and cluster analysis. Red numbers at the nodes represent approximately unbiased bootstrap values. g, Comparison between posterior forelimb digits and between neighbouring posterior hindlimb digits by Pearson's correlation coefficients.

To test this possibility we inspected the correlation coefficients between the posterior digits within each limb (Fig. 1g). In the early digit samples there is a strong negative correlation between the second and third wing digits but a weaker negative correlation between the second and third hindlimb digits and a positive correlation between the third and fourth hindlimb digits. In the late samples, the correlation between the second and third wing digits is still negative although weaker than for the early samples, whereas the correlations between the neighbouring posterior hindlimb digits are both positive. Together, these data indicate a much weaker differentiation between neighbouring posterior hindlimb digits than that between the posterior forelimb digits. We also performed differential expression analysis showing that the differences between posterior forelimb digits are larger than the difference between neighbouring posterior hindlimb digits (Supplementary Fig. 8). Overall, these results suggest that the second and third wing digits diverged and may have acquired derived digit identities during theropod evolution, making it difficult to find corresponding digit identities in the hindlimb.

To discover genes that specifically contribute to the second and third wing digit identities, we performed differential expression analysis of the mRNA-seq data between samples LFb and LFc. We found two genes, *Tbx3* and *Socs2*, with high expression in sample LFc (Supplementary Fig. 9 and Fig. 3a). To our knowledge no studies have been published indicating a role for *Socs2* in limb development. ISH confirms its strong expression in the third forelimb digit to the exclusion of all other digits in forelimb and hindlimb (Fig. 3b–g). Recently it has been shown that the third forelimb digit has a unique mode of

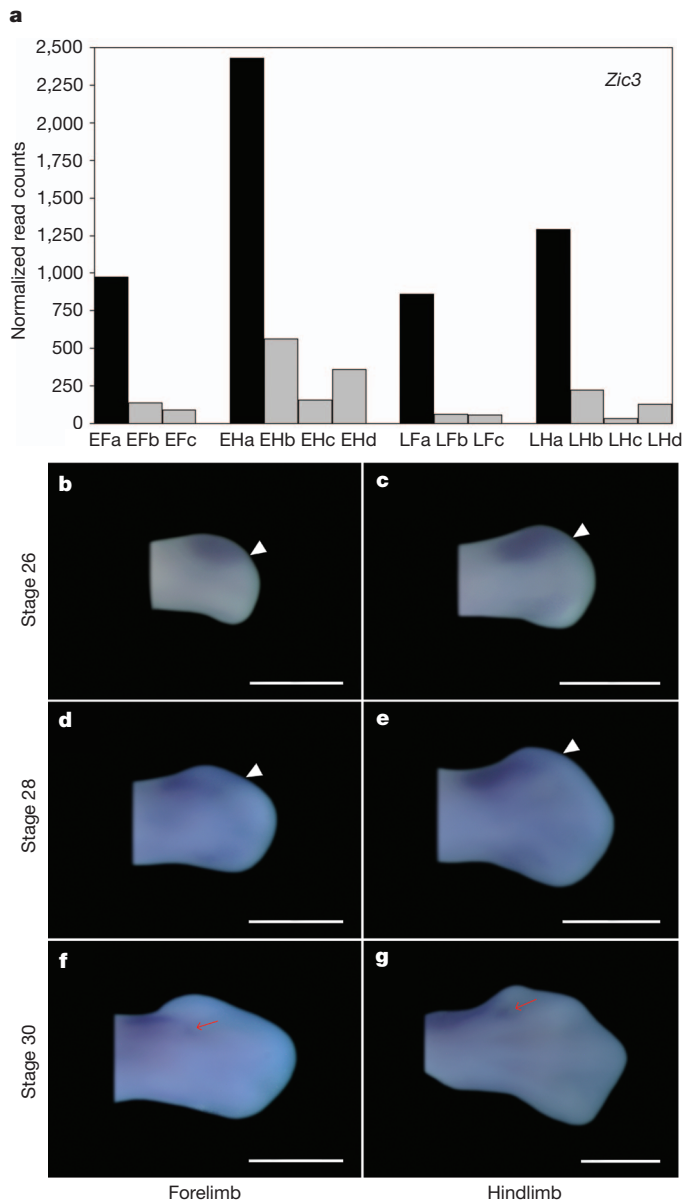


Figure 2 | Expression of *Zic3*. **a**, Expression pattern of *Zic3* in mRNA-seq samples. The y axis is the trimmed mean of M-values (TMM)³⁰ normalized mRNA-seq read-count value. **b–g**, Expression of *Zic3* in forelimb and hindlimb buds visualized by ISH. **b, c**, Stage 26; **d, e**, stage 28; **f, g**, stage 30. The dorsal view of the autopod is shown with anterior up and distal to the right. The white arrowheads in **b–e** indicate the expression boundary between first and second digit. The red arrows in **f, g** indicate the proximal expression of *Zic3*. Scale bar, 1 mm.

development in birds⁸. This, combined with our gene expression survey, supports the idea that the third wing digit has a unique derived identity in birds.

The data presented here reveal two patterns. First, there is a strong signal that identifies the anterior-most digits in the forelimb and hindlimb as homologous, in spite of the fact that they develop in different embryological positions. The second pattern is that no clear correspondence was found between the posterior forelimb digits and the posterior hindlimb digits. A comparison of digits within each limb shows that the posterior forelimb digits are more strongly differentiated than the posterior hindlimb digits, and forelimb digit III exhibits a unique expression of *Socs2* (Fig. 3d, f and Supplementary Fig. 10). These results provide evidence that the second and third wing digits

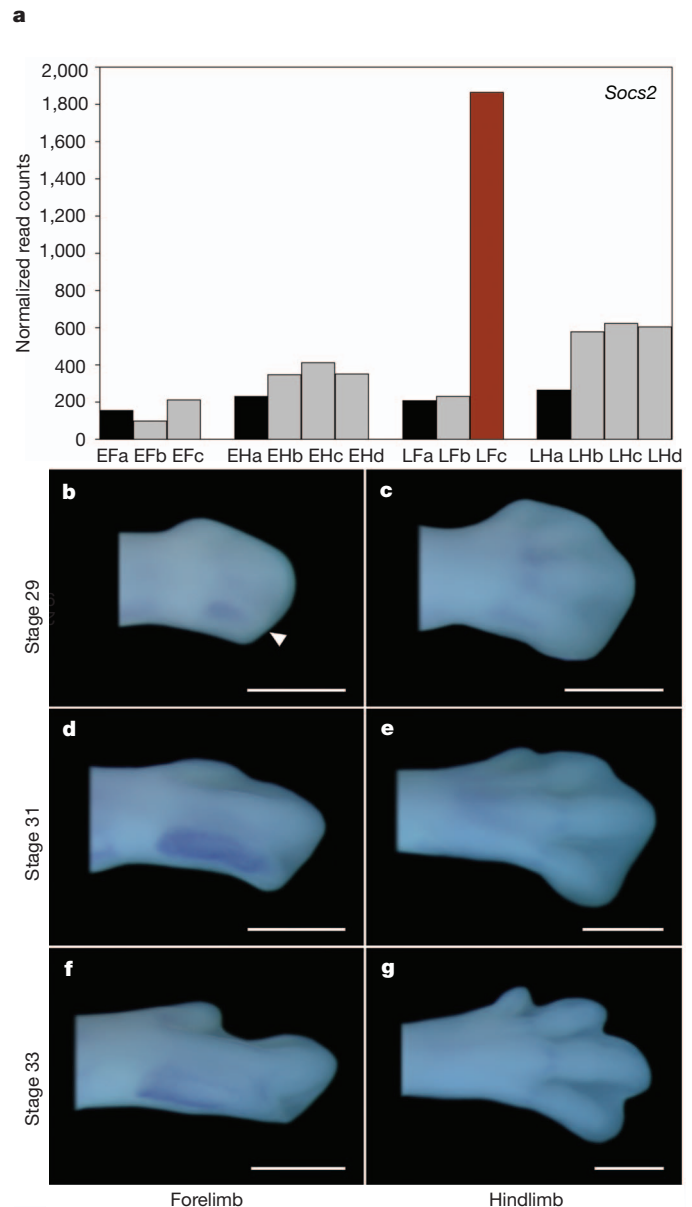


Figure 3 | Expression of *Socs2*. **a**, Expression pattern of *Socs2* in mRNA-seq samples. The y axis is the TMM normalized mRNA-seq read-count value. **b–g**, Expression of *Socs2* in forelimb and hindlimb digits visualized by ISH. **b, c**, Stage 29; **d, e**, stage 31; **f, g**, stage 33. Orientation of limb buds as in Fig. 2. The white arrowhead in **b** points at the expression domain at the third digit. Scale bar, 1 mm.

are developmentally distinct from the second to fourth hindlimb digits, indicating that digit individualization can be dynamic and evolutionarily transient.

METHODS SUMMARY

Fourteen samples and two pools of twenty individuals from embryonic chicken digits were used to perform mRNA-seq on an Illumina Genome Analyzer. Digital primordia with their closely posterior interdigital mesoderm were dissected separately from the forelimbs and hindlimbs between stages 28 and 29 (stage 28/29; Fig. 1a, b). Digital rays without the interdigital tissues were collected from the forelimbs and hindlimbs at stage 31 (Fig. 1c, d). mRNA-seq data were used for clustering^{17,27}, correlation analysis, differential gene expression²⁸ and Gene Ontology analysis. For ISH, we followed a previously published protocol²⁹. All the probes are labelled with digoxigenin (Roche, catalogue no. 11277073910) and hybridized with the chicken embryos at 70 °C.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 21 March; accepted 26 July 2011.

Published online 4 September 2011.

1. Novershtern, N. *et al.* Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell* **144**, 296–309 (2011).
2. Cherbas, L. *et al.* The transcriptional diversity of 25 *Drosophila* cell lines. *Genome Res.* **21**, 301–314 (2011).
3. Alizadeh, A. A. *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503–511 (2000).
4. Wagner, G. P. The developmental evolution of avian digit homology: an update. *Theory Biosci.* **124**, 165–183 (2005).
5. Young, R. L., Bever, G. S., Wang, Z. & Wagner, G. P. Identity of the avian wing digits: problems resolved and unsolved. *Dev. Dyn.* **240**, 1042–1053 (2011).
6. Burke, A. C. & Feduccia, A. Developmental patterns and the identification of homologies in the avian hand. *Science* **278**, 666–668 (1997).
7. Wagner, G. P. & Gauthier, J. A. 1,2,3 = 2,3,4: a solution to the problem of the homology of the digits in the avian hand. *Proc. Natl Acad. Sci. USA* **96**, 5111–5116 (1999).
8. Tamura, K. *et al.* Embryological evidence identifies wing digits in birds as digits 1, 2, and 3. *Science* **331**, 753–757 (2011).
9. Burke, A. C., Nelson, C. E., Morgan, B. A. & Tabin, C. Hox genes and the evolution of vertebrate axial morphology. *Development* **121**, 333–346 (1995).
10. Mansfield, J. H. & Abzhanov, A. Hox expression in the American alligator and evolution of archosaurian axial patterning. *J. Exp. Zool. B* **314**, 629–644 (2010).
11. Averof, M. & Akam, M. Hox genes and the diversification of insect and crustacean body plans. *Nature* **376**, 420–423 (1995).
12. Arendt, D. The evolution of cell types in animals: emerging principles from molecular studies. *Nature Rev. Genet.* **9**, 868–882 (2008).
13. Sugino, K. *et al.* Molecular taxonomy of major neuronal classes in the adult mouse forebrain. *Nature Neurosci.* **9**, 99–107 (2006).
14. Palmer, C., Diehn, M., Alizadeh, A. A. & Brown, P. O. Cell-type specific gene expression profiles of leukocytes in human peripheral blood. *BMC Genomics* **7**, 115 (2006).
15. Wagner, G. P. The developmental genetics of homology. *Nature Rev. Genet.* **8**, 473–479 (2007).
16. Young, R. L. & Wagner, G. P. Why ontogenetic homology criteria can be misleading: lessons from digit identity transformations. *J. Exp. Zool. B* **316B**, 165–170 (2011).
17. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
18. Welten, M. C., Verbeek, F. J., Meijer, A. H. & Richardson, M. K. Gene expression and digit homology in the chicken embryo wing. *Evol. Dev.* **7**, 18–28 (2005).
19. Tickle, C. Making digit patterns in the vertebrate limb. *Nature Rev. Mol. Cell Biol.* **7**, 45–53 (2006).
20. Vargas, A. O. & Fallon, J. F. Birds have dinosaur wings: The molecular evidence. *J. Exp. Zool. B* **304B**, 86–90 (2005).
21. Vargas, A. O. *et al.* The evolution of HoxD-11 expression in the bird wing: insights from *Alligator mississippiensis*. *PLoS ONE* **3**, e3325 (2008).
22. Uejima, A. *et al.* Anterior shift in gene expression precedes anteriormost digit formation in amniote limbs. *Dev. Growth Differ.* **52**, 223–234 (2010).
23. Montavon, T., Le Garrec, J. F., Kerszberg, M. & Duboule, D. Modeling Hox gene regulation in digits: reverse collinearity and the molecular origin of thumbness. *Genes Dev.* **22**, 346–359 (2008).
24. Firulli, B. A. *et al.* Altered Twist1 and Hand2 dimerization is associated with Saethre-Chotzen syndrome and limb abnormalities. *Nature Genet.* **37**, 373–381 (2005).
25. Zhu, L., Zhou, G., Poole, S. & Belmont, J. W. Characterization of the interactions of human ZIC3 mutants with GLI3. *Hum. Mutat.* **29**, 99–105 (2008).
26. Tzchori, I. *et al.* LIM homeobox transcription factors integrate signaling events that control three-dimensional limb patterning and growth. *Development* **136**, 1375–1385 (2009).
27. Suzuki, R. & Shimodaira, H. Pvcust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**, 1540–1542 (2006).
28. Wang, L. K. *et al.* DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* **26**, 136–138 (2010).
29. Hargrave, M., Bowles, J. & Koopman, P. *In situ* hybridization of whole-mount embryos. *Methods Mol. Biol.* **326**, 103–113 (2006).
30. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors thank J. Noonan for discussions on this project, K. Cooper for experimental assistance and N. Carriero for read-mapping assistance and C. Tabin for providing us with the Hoxd12 probe. The authors are also grateful for the technical support for this project by the Yale Center for Genomic Analysis. The financial support by the Yale Science Development Fund is gratefully acknowledged.

Author Contributions Z.W. performed the experiments and data analysis and participated in design of the study. R.L.Y., H.X. and G.P.W. participated in data analysis. G.P.W. conceived and designed the study and supervised the work. All authors discussed the results and made substantial contributions to the manuscript.

Author Information All mRNA-seq data are deposited in the Gene Expression Omnibus under accession number GSE28156. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to G.P.W. (gunter.wagner@yale.edu).

METHODS

Sample collection. Fertilized chicken eggs were obtained from Charles River Laboratories International and incubated between 37.5 and 39 °C at high relative humidity (>50%). To obtain sufficient RNA for transcriptome sequencing, digit samples were collected and pooled from 20 individuals. Previous studies show that digit identity is regulated by the closely posterior interdigital mesoderm and that digit identities of digit I/II and III are not fixed until stages 30 and 29 in the chicken hindlimb^{31,32}. Thus, we use digital primordia with their closely posterior interdigital mesoderm to compose a sample for mRNA-seq at stage 28/29, and only the digital ray as a sample at stage 31 (Fig. 1a–d). Collections were performed at two time points during egg incubation to obtain stage 28/29 and stage 31 embryos. Staging is based on Hamburger and Hamilton³³. Digital primordia with their closely posterior interdigital mesoderm were dissected separately from the forelimb and hindlimb between stages 28 and 29 (stage 28/29; Fig. 1a, b). Digital rays without interdigital tissues were collected from the forelimb and hindlimb at stage 31 (Fig. 1c, d). In both cases, we removed the tissue anterior to the first digital ray of the forelimb. This tissue is thought to be the residual condensation of the ancestral digit I¹⁸.

mRNA-seq. Total RNA was extracted using the RNeasy Mini kit (Qiagen), and treated with DNase I using RNase-Free DNase set (Qiagen). Extracted total RNA of the samples was submitted to Yale Center for Genome Analysis for mRNA sequencing (mRNA-seq) using an Illumina Genome Analyzer. mRNA-Seq 8-Sample Prep kit (Illumina), Standard Cluster Generation kit (Illumina) and Illumina Sequencing kit (Illumina) were used for constructing libraries for single-read sequencing, cluster amplification and final sequencing-by-synthesis. Two flow cells, each containing 8 lanes and 8 samples including the Phi X174 control, were used. Image analysis, base calling, extraction of 34-bp reads, and read counting were performed using the Illumina pipeline.

Sequence alignment, gene counting and normalization. Reads from mRNA-seq data were aligned to Ensembl chicken genome (*Gallus gallus* WASHUC2) using the GERALD module of Illumina CASAVA v1.6 software. After alignment, raw gene counts (gene expression levels) were counted by Illumina CASAVA v1.6 software, and then normalized using the TMM method, which takes RNA composition bias into account, using the edgeR package^{17,30}. The normalization factors were calculated by edgeR^{17,30} for each library and then the normalized read counts were calculated by the following formula: normalized count = raw count ÷ (library size × normalization factor) × 10⁶. The normalized count was directly used for subsequent analyses (see below), and was also divided by the cDNA length to adjust for gene length and then used in the same analysis.

Differential expression analysis. DESeq²⁸ packages were used to identify differentially expressed genes between digit I (EFa, LFa, EHa and LH; Fig. 1a–d) and all the other digits (EFb, EFc, LFb, LFc, Ehb, Ehc, Ehd, Lhb, Lhc and Lhd; Fig. 1a–d), and between Lfb and Lfc samples (Fig. 1c, d). Function DEGexp was performed to identify differentially expressed genes and the R scripts for using this function are provided in the User's Guide of DESeq²⁸. DESeq²⁸ was also used to identify differentially expressed genes between forelimb (EFa, EFb, EFc, LFa, Lfb and Lfc; Fig. 1a–d) and hindlimb digits (EHa, Ehb, Ehc, Ehd, LHa, Lhb, Lhc and Lhd; Fig. 1a–d), and these genes were removed from the input data for the MDS plot (see below). Owing to the underestimate of the true variability (resulting from pooling individuals), more differentially expressed genes may be removed, which actually makes the downstream approach (MDS plot) conservative. A gene is considered significantly differentially expressed when the *Q*-value (adjusted *P* value for multiple testing) is lower than 0.05 (ref. 34).

MDS plot. The MDS plot is conducted by the edgeR package¹⁷. The input data are normalized read counts of 5,813 genes after removing genes that are significantly differentially expressed between forelimb digits and hindlimb digits. The distance between each pair of samples in the output figure is the square root of the common dispersion for the top 100 genes that best distinguish that pair of samples¹⁷. These top 100 genes are selected according to the tagwise dispersion of all the samples¹⁷.

Cluster analysis and heat map of correlation. We produced two sets of data from the normalized data by different methods. The two data sets are separately used for hierarchical clustering and calculating either Pearson's correlation coefficient or Spearman's rank correlation coefficient. The first method considers the mRNA-seq data to follow a Poisson distribution and calculates the square root values of the normalized data (the canonical variance-stabilizing transformation for Poisson). To calculate the input data, the 14 samples are divided into four groups: group 1

(EFa, EFb and EFc), group 2 (EHa, Ehb, Ehc and Ehd), group 3 (LFa, Lfb and Lfc) and group 4 (LHa, Lhb, Lhc and Lhd) depending on developmental stages and forelimb/hindlimb. A distance is calculated by subtracting the mean of square root values of a group (for example, group 1: all the forelimb digits at stage 28/29) from each square root value of a sample (for example, EFa). The distance matrix is used for clustering and calculating the correlation coefficient between samples.

The second method considers that the normal distribution, with mean λ and variance λ (standard deviation: the square root of λ), is an approximation to the Poisson distribution because of large values of λ (in this study, $\lambda \approx 700$). Thus, a standard score (also called *z*-value, *z*-score), which indicates how many standard deviations a datum is above or below the mean, is used for clustering and calculating Pearson's correlation coefficient between samples.

A heat map of correlation coefficients between samples was created by the R Lattice package. Clustering is conducted by pvclust, an R package for hierarchical clustering with *P* values²⁷. After bootstrap re-sampling (10,000 iterations), approximately unbiased (AU) *P* value is provided in the output figure. AU *P* value, which is calculated by multiscale bootstrap re-sampling, is a better approximation to unbiased *P* value than bootstrap probability value calculated by ordinary bootstrap re-sampling. Clusters are formed using the average dissimilarity between the samples. An AU *P* value higher than 95% indicates that the cluster is highly supported²⁷.

Gene Ontology analysis. Gene Ontology analysis is performed by the GOseq package³⁵, which takes gene length bias into account. We first identify the significantly differentially expressed genes that are driving the clustering by edgeR package¹⁷ and then identify significantly enriched GO categories using a 0.05 cutoff for the false discovery rate. Default parameters and methods in the packages are used.

In situ hybridization. Total RNA was extracted from chicken embryonic limb buds using the RNeasy Mini kit (Qiagen), and treated with DNase I using RNase-Free DNase set (Qiagen). cDNA was synthesized from the total RNA using the High Capacity cDNA Reverse Transcription kit (Applied Biosystems). The whole coding sequence of *Hand2* was amplified using a pair of specifically designed primers (forward primer, 5'-GCGGCGATGAGTCTTGTG-3'; reverse primer, 5'-CTCACTGCTTGAGCTCCAGC-3'; product, 658 bp). Partial coding sequence followed by partial 3' UTR of *Zic3* was amplified using a pair of primers previously reported but without the additional T7 RNA polymerase promoter sequence (forward primer, 5'-CAGCAAGGACTCCACGAAAAC-3'; reverse primer, 5'-CGACCCCATCAGATGAGAAT-3'; product, 723 bp)³⁵. Most of the coding sequence of *Lhx9* (914 bp) was amplified using a pair of primers previously reported³⁷. Most of the coding sequence followed by partial 3' UTR of *Socs2* was amplified using a pair of specifically designed primers (forward primer, 5'-CGTTGCCGAAGCCAAGGAGA-3'; reverse primer, 5'-AGGGATGCGAG CGGGGATAA-3'; product, 624 bp). PCR products were purified using the QIAquick Gel Extraction kit (Qiagen) and cloned into pGEM-T vector (Promega). All the antisense probes were prepared by linearizing the plasmid with NcoI and transcribing with SP6 polymerase, whereas the sense probes were prepared by linearizing the plasmid with SpeI and transcribing with T7 polymerase. The probe of *Hoxd12* was provided by C. Tabin³⁸. All of the probes are labelled with digoxigenin (Roche) and hybridized with the chicken embryos at 70 °C. The procedure of ISH followed a previously published protocol³².

- Dahn, R. D. & Fallon, J. F. Interdigital regulation of digit identity and homeotic transformation by modulated BMP signaling. *Science* **289**, 438–441 (2000).
- Suzuki, T., Hasso, S. M. & Fallon, J. F. Unique SMAD1/5/8 activity at the phalanx-forming region determines digit identity. *Proc. Natl Acad. Sci. USA* **105**, 4185–4190 (2008).
- Hamburger, V. & Hamilton, H. L. A series of normal stages in the development of the chick embryo. *J. Morphol.* **88**, 49–92 (1951).
- Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
- Young, M. D., Wakefield, M. J., Smyth, G. K. & Oshlack, A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* **11**, R14 (2010).
- McMahon, A. R. & Merzdorf, C. S. Expression of the *zic1*, *zic2*, *zic3*, and *zic4* genes in early chick embryos. *BMC Res. Notes* **3**, 167 (2010).
- Abellán, A. *et al.* Olfactory and amygdalar structures of the chicken ventral pallidum based on the combinatorial expression patterns of LIM and other developmental regulatory genes. *J. Comp. Neurol.* **516**, 166–186 (2009).
- Nelson, C. E. *et al.* Analysis of Hox gene expression in the chick limb bud. *Development* **122**, 1449–1466 (1996).

The genome of the green anole lizard and a comparative analysis with birds and mammals

Jessica Alföldi^{1*}, Federica Di Palma^{1*}, Manfred Grabherr¹, Christina Williams², Lesheng Kong³, Evan Mauceli¹, Pamela Russell¹, Craig B. Lowe⁴, Richard E. Glor⁵, Jacob D. Jaffe¹, David A. Ray⁶, Stephane Boissinot⁷, Andrew M. Shedlock⁸, Christopher Botka⁹, Todd A. Castoe¹⁰, John K. Colbourne¹¹, Matthew K. Fujita^{3,12}, Ricardo Godinez Moreno¹², Boudewijn F. ten Hallers¹³, David Haussler¹⁴, Andreas Heger³, David Heiman¹, Daniel E. Janes¹², Jeremy Johnson¹, Pieter J. de Jong¹³, Maxim Y. Koriabine¹³, Marcia Lara¹, Peter A. Novick¹⁵, Chris L. Organ¹², Sally E. Peach¹, Steven Poe¹⁶, David D. Pollock^{10,17}, Kevin de Queiroz¹⁸, Thomas Sanger¹², Steve Searle¹⁹, Jeremy D. Smith⁶, Zachary Smith¹¹, Ross Swofford¹, Jason Turner-Maier¹, Juli Wade²⁰, Sarah Young¹, Amonida Zadissa¹⁹, Scott V. Edwards¹², Travis C. Glenn²¹, Christopher J. Schneider²², Jonathan B. Losos¹², Eric S. Lander¹, Matthew Breen^{2,23,24}, Chris P. Ponting³ & Kerstin Lindblad-Toh^{1,25}

The evolution of the amniotic egg was one of the great evolutionary innovations in the history of life, freeing vertebrates from an obligatory connection to water and thus permitting the conquest of terrestrial environments¹. Among amniotes, genome sequences are available for mammals and birds^{2–4}, but not for non-avian reptiles. Here we report the genome sequence of the North American green anole lizard, *Anolis carolinensis*. We find that *A. carolinensis* microchromosomes are highly syntenic with chicken microchromosomes, yet do not exhibit the high GC and low repeat content that are characteristic of avian microchromosomes². Also, *A. carolinensis* mobile elements are very young and diverse—more so than in any other sequenced amniote genome. The GC content of this lizard genome is also unusual in its homogeneity, unlike the regionally variable GC content found in mammals and birds⁵. We describe and assign sequence to the previously unknown *A. carolinensis* X chromosome. Comparative gene analysis shows that amniote egg proteins have evolved significantly more rapidly than other proteins. An anole phylogeny resolves basal branches to illuminate the history of their repeated adaptive radiations.

The amniote lineage divided into the ancestral lineages of mammals and reptiles ~320 million years ago. Today, the surviving members of those lineages are mammals, comprising ~4,500 species, and reptiles, containing ~17,000 species. Within the reptiles, the two major clades diverged ~280 million years ago: the lepidosaurs, which contains lizards (including snakes) and the tuatara; and the archosaurs, containing crocodilians and birds (the position of turtles remains unclear)⁶. For simplicity, we will refer here to lepidosaurs as lizards (Fig. 1).

The study of the major genomic events that accompanied the transition to a fully terrestrial life cycle has been assisted by the sequencing of several mammal (K.L.-T. *et al.*, manuscript submitted) and three bird genomes^{2–4}. The genome of the lizard *A. carolinensis* thus fills an important gap in the coverage of amniotes, splitting the long branch between mammals and birds and allowing more robust evolutionary analysis of amniote genomes.

For instance, almost all reptilian genomes contain microchromosomes, but these have only been studied at a sequence level in birds^{2,7}, raising the question as to whether the avian microchromosomes' peculiar sequence features are universal across reptilian microchromosomes⁸. Another example is the study of sex chromosome evolution. Nearly all placental and marsupial mammals share homologous sex chromosomes (XY)⁹ and all birds share ZW sex chromosomes. However, lizards exhibit either genetic or temperature-dependent sex determination¹⁰. Characterization of lizard sex chromosomes would allow the study of previously unknown sex chromosomes and comparison of independent sex chromosome systems in closely related species.

Anolis lizards comprise a diverse clade of ~400 described species distributed throughout the Neotropics. These lizards have radiated, often convergently, into a variety of ecological niches with attendant morphological adaptations, providing one of the best examples of adaptive radiation. In particular, their diversification into multiple replicate niches on diverse Caribbean islands via interspecific competition and natural selection has been documented in detail¹¹. *A. carolinensis* is the only anole native to the USA and can be found from Florida and Texas up to North Carolina. We chose this species for genome sequencing because it is widely used as a reptile model for experimental ecology, behaviour, physiology, endocrinology, epizootics and, increasingly, genomics.

The green anole genome was sequenced and assembled (AnoCar 2.0) using DNA from a female *A. carolinensis* lizard (Supplementary Tables 1–4). Fluorescence *in situ* hybridization (FISH) of 405 bacterial artificial chromosome (BAC) clones (from a male) allowed the assembly scaffolds to be anchored to chromosomes (Supplementary Table 5 and Supplementary Fig. 1). The *A. carolinensis* genome has been reported to have a karyotype of $n = 18$ chromosomes, comprising six pairs of large macrochromosomes and 12 pairs of small microchromosomes¹². The draft genome sequence is 1.78 Gb in size (see Supplementary Table 3 for assembly statistics) and represents an intermediate

¹Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. ²Department of Molecular Biomedical Sciences, College of Veterinary Medicine, North Carolina State University, Raleigh, North Carolina 27606, USA. ³MRC Functional Genomics Unit, University of Oxford, Department of Physiology, Anatomy and Genetics, Oxford OX1 3QX, UK. ⁴Stanford University School of Medicine, Department of Developmental Biology, Stanford, California 94305, USA. ⁵University of Rochester, Rochester, New York 14607, USA. ⁶Department of Biochemistry, Molecular Biology, Entomology and Plant Pathology, Mississippi State University, Mississippi State, Mississippi 39762, USA. ⁷Department of Biology, Queens College, the City University of New York, New York, New York 11367, USA. ⁸Biology Department and Graduate Program in Marine Genomics, College of Charleston, Charleston, South Carolina 29424, USA. ⁹Harvard Medical School, Boston, Massachusetts 02115, USA. ¹⁰Department of Biochemistry and Molecular Genetics, University of Colorado School of Medicine, Aurora, Colorado 80045, USA. ¹¹The Center for Genomics and Bioinformatics, Indiana University, Bloomington, Indiana 47405, USA. ¹²Department of Organismic and Evolutionary Biology and Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138, USA. ¹³Children's Hospital Oakland Research Institute, Oakland, California 94609, USA. ¹⁴Center for Biomolecular Science and Engineering, University of California, Santa Cruz, California 95064, USA. ¹⁵Department of Biological Sciences and Geology, Queensborough Community College, Bayside, New York 11364, USA. ¹⁶Department of Biology, University of New Mexico, Albuquerque, New Mexico 87131, USA. ¹⁷Program in Computational Bioscience, University of Colorado School of Medicine, Aurora, Colorado 80045, USA. ¹⁸Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, District of Columbia 20560, USA. ¹⁹Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK. ²⁰Departments of Psychology and Zoology, Program in Neuroscience, Michigan State University, East Lansing, Michigan 48824, USA. ²¹Department of Environmental Health Science and Georgia Genomics Facility, University of Georgia, Athens, Georgia 30602, USA. ²²Biology Department, Boston University, Boston, Massachusetts 02215, USA. ²³Center for Comparative Medicine and Translational Research, North Carolina State University, Raleigh, North Carolina 27695, USA. ²⁴University of North Carolina Lineberger Comprehensive Cancer Center, Chapel Hill, North Carolina 27514, USA. ²⁵Science for Life Laboratory Uppsala, Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala 751 23, Sweden.

*These authors contributed equally to this work.

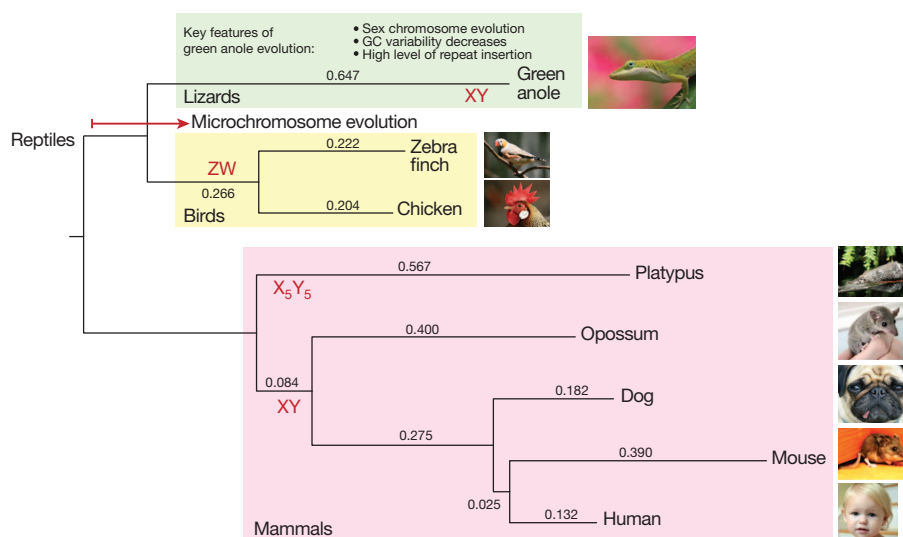


Figure 1 | Amniote phylogeny based on protein synonymous sites showing major features of amniote evolution. Major characteristics of lizard evolution including homogenization of GC content, high sex chromosome turnover and

high levels of repeat insertion are featured. Sex chromosome inventions are indicated in red. Branch length is proportional to dS (the synonymous substitution rate); dS of each branch is indicated above the line.

between genome assemblies of birds (0.9–1.3 Gb) and mammals (2.0–3.6 Gb).

We find that few chromosomal rearrangements occurred in the 280 million years since anole and chicken diverged, as had been hinted at by previous comparisons using *Xenopus* and chicken¹³. There are 259 syntenic blocks (defined as consecutive syntenic anchors that are consistent in order, orientation and spacing, at a resolution of 1 Mb) between lizard and chicken (Supplementary Table 6 and Supplementary Fig. 2). Interestingly, 19 out of 22 anchored chicken chromosomes are each syntenic to a single *A. carolinensis* chromosome over their entire lengths (Fig. 2a); by contrast, only 6 (of 23) human chromosomes are syntenic to a single opossum chromosome over their entire lengths, even though the species diverged only 148 million years ago¹⁴. Segmental duplications follow trends seen in other amniote genomes (Supplementary Note, Supplementary Table 7 and Supplementary Fig. 3).

Approximately 30% of the *A. carolinensis* genome is composed of mobile elements, which comprise a much wider variety of active repeat families than is seen for either bird² or mammalian¹⁵ genomes. The most active classes are long interspersed (LINE) elements (27%) and short interspersed (SINE) elements (16%)¹⁶ (Supplementary Table 8). The majority of LINE repeats belong to five groups (L1, L2, CR1, RTE and R4) and seem to be recent insertions based on their sequence similarity (divergence ranges from 0.00–0.76%; ref. 17). This contrasts with observations of mammalian genomes, where only a single family of LINEs—L1—has predominated over tens of millions of years. The DNA transposons comprise at least 68 families belonging to five superfamilies: hAT, Chapaev, Maverick, Tc/Mariner and Helitron¹⁸. As with retrotransposons, the majority of DNA transposon families seem to be relatively young in contrast to the extremely few recently active DNA transposons found in other amniote genomes (Supplementary Table 9). Overall, *A. carolinensis* mobile elements feature significantly higher GC content (43.5%, $P < 10^{-20}$) than the genome-wide average of 40.3%. In addition to mobile elements, *A. carolinensis* exhibits a high density (3.5%) of tandem repeats, with length and frequency distributions similar to those of human microsatellite DNA¹⁵. We now know that amniote genomes come in at least three types: mammalian genomes are enriched for L1 elements and have a high degree of mobile element accumulation, bird genomes are repeat poor with very little mobile element activity, while the lizard genome contains an extremely wide diversity of active mobile element families but has a low rate of accumulation, which is reminiscent of the mobile element profile of teleostean fishes¹⁹.

Most reptile genomes contain microchromosomes, but the numbers vary among species; the *A. carolinensis* genome contains 12 pairs of microchromosomes¹², whereas the chicken genome contains 28 pairs. Bird microchromosomes have very distinctive properties compared to bird macrochromosomes, such as higher GC and lower repeat contents², whereas lizard microchromosomes do not exhibit these features (Fig. 2b). Remarkably, all sequence anchored to microchromosomes in *A. carolinensis* also aligns to microchromosomes in the chicken genome, and all but one *A. carolinensis* microchromosome is syntenic to only a single corresponding chicken microchromosome (Fig. 2a). Microchromosomes conserved between *A. carolinensis* and chicken thus could have arisen in the reptile ancestor, whereas the remaining chicken microchromosomes could be derived in the bird lineage. Alternatively, the remaining chicken microchromosomes could have been present in the reptile ancestor but fused to form macrochromosomes in the lizard lineage.

The *A. carolinensis* genome has surprisingly little regional variation of GC content, substantially less than previously observed for birds and mammals; it is the only amniotic genome known whose nucleotide composition is as homogenous as the frog genome⁵ (Supplementary Figs 4 and 5). Figure 3 illustrates how local GC content is evolutionarily conserved between human chromosome 14 and chicken chromosome 5, but to a much lesser degree with *A. carolinensis* chromosome 1. As all sequenced amniote genomes other than *A. carolinensis* contain these homologous varying levels of GC content ('isochores')²⁰, the ancestral amniote GC heterogeneity is likely to have eroded towards homogeneity in this lizard's lineage. It has been proposed that isochores with high GC content are a consequence of higher rates of GC-biased gene conversion in regions of higher recombination². The greater GC homogeneity in the anole genome may thus reflect more uniform recombination rates, or else a substantially reduced bias towards GC during the resolution of gene conversion events in the *A. carolinensis* lineage (for a discussion, see ref. 5).

Both temperature-dependent sex determination and XY genetic sex determination have been found in Iguania¹⁰. Within the genus *Anolis*, there are species with heteromorphic XY chromosomes (including those with multiple X and Y chromosomes), and others with entirely homomorphic chromosomes¹². *A. carolinensis* is known to have genetic sex determination²¹, but the form of its sex chromosomes (ZW or XY) has thus far been unknown owing to a lack of obviously heteromorphic chromosomes.

In depth examination of male and female cells using FISH allowed us to identify the microchromosome previously designated as 'b' as the

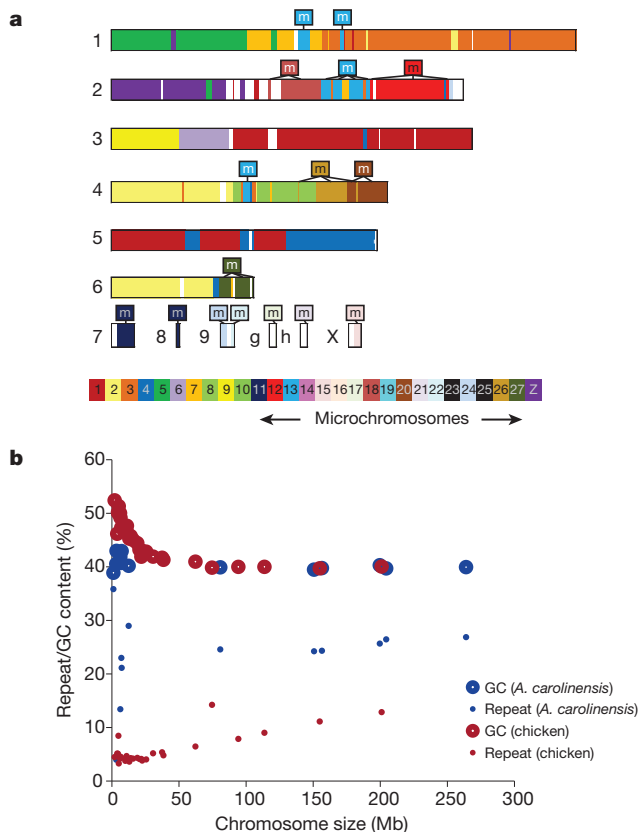


Figure 2 | *A. carolinensis*–chicken synteny map reveals synteny of reptile microchromosomes but dissimilar GC and repeat content. **a**, Very few rearrangements have occurred in the 280 million years since *A. carolinensis* and chicken diverged. *A. carolinensis* microchromosomes are exclusively syntenic to chicken microchromosomes. Horizontal coloured bars depict the six *A. carolinensis* macrochromosomes (1–6) and the six (of 12) *A. carolinensis* microchromosomes that have sequence anchored to them that is syntenic to the chicken genome (7, 8, 9, X, LGg, LGh). Chromosomes that could be ordered by size were assigned a number; the smaller microchromosomes that could not be distinguished by size were assigned a lowercase letter. Each colour corresponds to a different chicken chromosome as indicated in the key. Any part of an *A. carolinensis* chromosome that is syntenic to a chicken microchromosome is indicated by 'm'. **b**, Chicken microchromosomes have both higher GC content and lower repeat content than chicken macrochromosomes, whereas *A. carolinensis* chromosomes do not vary in GC or repeat content by chromosome size. Large circles designate the GC percentage of each chromosome in the chicken and lizard genomes with greater than 100 kb of sequence anchored to it. Small circles designate the percentage of the genome made up of repetitive sequence of each chromosome in the chicken (blue circles) and lizard (red circles) genomes.

A. carolinensis X chromosome; it is present in two copies in females and one in males. This chromosome is syntenic to chicken microchromosome 15. Eleven BACs assigned to two scaffolds, 154 (3.3 Mb) and chrUn0090 (1.8 Mb), hybridize via FISH to the p arms of the two X chromosomes in females, and hybridize to the p arm of the single X chromosome in males (Fig. 4 and Supplementary Fig. 1). *A. carolinensis* thereby shows a pattern representative of a male heterogametic system of genotypic sex determination. We have not identified the Y chromosome, but we hypothesize that *A. carolinensis* possesses both X and Y chromosomes, as both male and female cells contain the same number of chromosomes.

The 5.1 Mb of sequence assigned to the X chromosome contains 62 protein-coding genes (Supplementary Table 10); Gene Ontology (GO) terms associated with these genes show no significant enrichment. It is very likely that there is more X chromosome sequence that is currently labelled as unanchored scaffolds in the AnoCar 2.0 assembly. Identification of the *A. carolinensis* sex determination gene will require

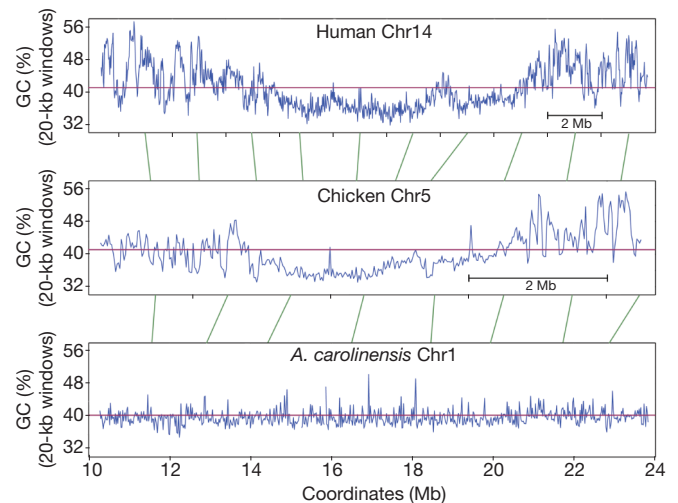


Figure 3 | The *A. carolinensis* genome lacks isochores. The *A. carolinensis* genome shows only very local variation in GC content, unlike the human and chicken genomes, which also show larger trends in GC variation, sometimes called isochores. Syntenic regions of human chromosome 14, chicken chromosome 5 and *A. carolinensis* chromosome 1 are shown. The human and chicken regions are inverted and rearranged to align with the *A. carolinensis* region. Blue lines depict GC percentage in 20-kb windows. The purple line designates the genome average. Green lines represent examples of syntenic anchors between the three genomes.

considerable functional biology, but we note that the chicken sex determination gene *DMRT1* is located on *A. carolinensis* chromosome 2 and that *SOX3* (the X chromosome paralogue of the therian mammal sex determination gene *SRY*) is located on an unanchored *A. carolinensis* scaffold; these genes are thus unlikely to be the *A. carolinensis* sex determination gene.

All ten *A. carolinensis* individuals (originating from South Carolina and Tennessee) used for FISH mapping showed large pericentromeric inversions in one or more of chromosomes 1–4, with no correlation between different chromosomal inversions or with the sex of the lizard (see Supplementary Note, Supplementary Table 11 and Supplementary Fig. 6).

A total of 17,472 protein-coding genes and 2,924 RNA genes were predicted from the *A. carolinensis* genome assembly (Ensembl release 56, September 2009). We built a phylogeny for all *A. carolinensis* genes and their homologues in eight other vertebrate species (human,

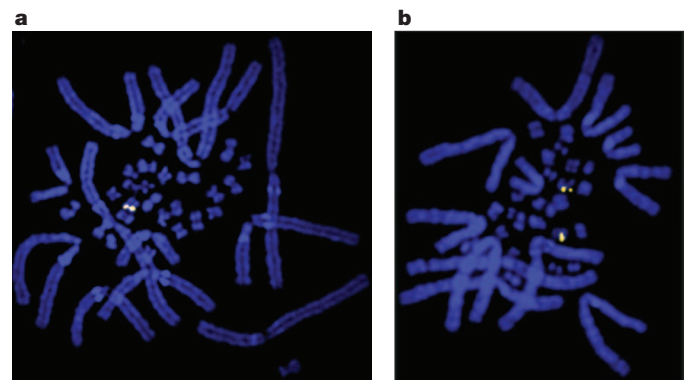


Figure 4 | The *A. carolinensis* genome contains a newly discovered X chromosome. **a**, **b**, The X chromosome, a microchromosome, is found in one copy in male *A. carolinensis* (**a**) and in two copies in females (**b**). The BAC 206M13 (CHORI-318 BAC library) is hybridized to the p arm of the X chromosome using FISH in both male and female metaphase spreads. 206M13 and ten other BACs showed this sex-specific pattern in cells derived from five male and five female individuals. Original magnification, $\times 1,000$.

mouse, dog, opossum, platypus, chicken, zebra finch and pufferfish), allowing us to identify a conservative set of 3,994 one-to-one orthologues, that is, genes that have not been duplicated or deleted in any of these vertebrates since their last common ancestor. These gene phylogenies were also used to identify genes that arose by duplication in the lizard lineage after the split with the avian lineage and, separately, those that were lost in the mammalian lineage after the mammal–reptile split (Fig. 1, Supplementary Note, Supplementary Fig. 7 and Supplementary Table 12).

We found 11 *A. carolinensis* opsin genes that have no mammalian orthologues (but have orthologues in invertebrates, fishes and frog), and thus seem to have been lost during mammalian evolution (Supplementary Table 13). The large repertoire of opsins may contribute to the excellent colour vision of anoles—including the ability to see in the ultraviolet range—and also may contribute to their hyperdiversity by allowing the evolution of diverse, species-specific colouration of the dewlap, which has an important role in sexual selection and species recognition¹¹. Similarly, olfactory receptor and β -keratin genes are highly duplicated in *A. carolinensis* (Supplementary Note and Supplementary Fig. 9).

Many reptiles, including green anoles, differ from placental mammals in being oviparous (laying eggs). Vivipary in placental mammals is a derived state, reflected in their loss of some egg-related genes. We used mass spectrometry to identify proteins present in the immature *A. carolinensis* egg, as most egg proteins are produced in the mother's body and then transported into the immature egg. We found that in contrast with mammals, reptiles have lineage-specific gene duplications, including in vitellogenins (VTGs), apovitellenin-1, ovomucin- α and three homologues of ovocalyxin-36, a chicken eggshell matrix protein.

Our results show rapid evolution of egg protein genes among amniotes. Specifically, we found proteins from 276 *A. carolinensis* genes in immature *A. carolinensis* eggs (Supplementary Tables 14 and 15), of which only 50 have been confirmed to be present in chicken eggs by mass spectrometry^{22,23}. These genes include VTGs, a lysozyme, vitelline membrane outer layer protein 1 (*VMO1*) paralogues, protease inhibitors, natterin and nothepsin. By aligning genes that are one-to-one orthologues in *A. carolinensis* and chicken, we found that egg proteins evolve significantly more rapidly than non-egg proteins (mean dN/dS values (ratio of the rate of non-synonymous substitutions to the rate of synonymous substitutions) of 0.186 and 0.135, respectively; $P = 1.2 \times 10^{-5}$), which reflects reduced purifying selection and/or more frequent episodes of adaptive evolution.

Using multiple vertebrate genome sequences, we identified three *VMO1* paralogues (which we name α , β and γ) that we infer to have been present in the last common ancestor of all reptiles and mammals. Whereas at least one of *VMO1*- α , *VMO1*- β and *VMO1*- γ has been lost in all other amniote genomes, the *A. carolinensis* genome contains representatives of all three paralogues. Moreover, the *A. carolinensis*-specific *VMO1*- α family has grown to 13 members and has experienced positive selection of amino acid substitutions within a negatively charged, probably substrate-binding cavity; changes that, presumably, modify its lysozyme-like transferase activity (Supplementary Note, Supplementary Fig. 8 and Supplementary Tables 16 and 17).

The extensive and active repeat repertoire of *A. carolinensis* has allowed us to discover the origin of several mammalian conserved elements. Through the process of exaptation (a major change in function of a sequence during evolution), certain mobile elements that were active in the amniote ancestor have become conserved, and presumably functional, in mammals, while remaining active mobile elements in *A. carolinensis*. The origin of these conserved mammalian sequences in mobile elements was not recognizable without comparison to a distant and repeat-rich genome sequence²⁴. We identified 96 such exapted elements (see Supplementary Table 18) in the human genome tracing back to mobile elements present in the amniote ancestor that are still present in *A. carolinensis*, particularly the CR1, L2 and gypsy families.

Although most exapted elements are non-coding and probably serve a regulatory function, we also identified a protein-coding exon that was exapted from an L2-like LINE, now constituting exon 2 in a mammal-specific N-terminal region of the *MIER1* (mesoderm induction early response 1) protein. This exon is highly conserved across 29 mammals and therefore probably represents a mammalian innovation since the amniote ancestor.

GO terms associated with the transcription start site closest to each exapted element in the human genome show enrichment for neurodevelopmental genes (see Methods), with “ephrin receptor binding”, “nervous system development” and “synaptic transmission” being strongly enriched (all P values $< 5 \times 10^{-3}$). These enrichments are consistent with adaptive changes in neurodevelopment occurring during the emergence of mammals.

Anolis lizards are a textbook case of adaptive radiation, having diversified independently on each island in the Greater Antilles and throughout the Neotropics, producing a wide variety of ecologically and morphologically differentiated species, with as many as 15 found at a single locality¹¹. Although anoles are widely used as a model system for phylogenetic comparative studies, it has been difficult to determine the evolutionary relationships among major anole clades owing to rapid evolutionary radiations associated with access to new dimensions of ecological opportunity. Successfully resolving the relatively short branching events associated with such a radiation requires a wealth of data from loci evolving at an appropriate rate.

We used the genome sequence of *A. carolinensis* to develop a new phylogenomic data set comprised of 20 kb of sequence data sampled from across the genomes of 93 species of anoles (Supplementary Tables 19 and 20). Analyses of this data set infer a well-supported phylogeny that reinforces and clarifies the adaptive and biogeographic history of anoles (Fig. 5, details in Supplementary Fig. 10). First, our phylogenomic analysis reaffirms previous molecular and morphological studies indicating that similar anole habitat specialists have

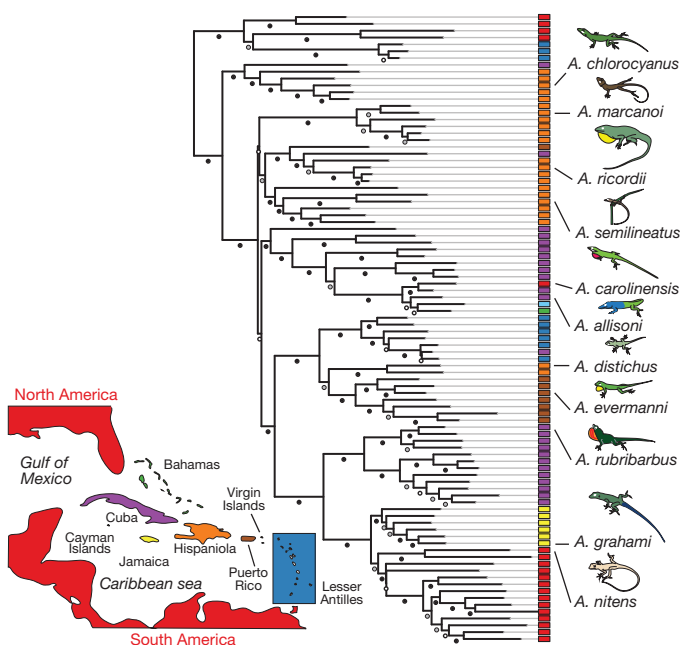


Figure 5 | A phylogeny of 93 *Anolis* species clarifies the biogeographic history of anoles. *Anolis* ecomorphs derive from convergent evolution and not from frequent inter-island migration. Using conserved primer pairs distributed across the genome of *A. carolinensis*, we obtain sequences from 46 genomically diverse loci evolving at a range of evolutionary rates and representing both protein-coding and non-coding regions. Maximum likelihood analyses of this new data set of 20 kb aligned nucleotides infer nearly all previously established anole relationships while also partially resolving the basal relationships that have plagued previous studies. Open circles indicate bootstrap (bs) values < 70 ; grey-shaded circles, $70 < bs < 95$; filled circles, $bs > 95$.

evolved independently on each of the four large Greater Antillean islands. Second, our analyses suggest a complex biogeographic scenario involving a limited number of dispersal events between islands and extensive *in situ* diversification within islands. The closest relatives of *Anolis* occur on the mainland and the phylogeny confirms the existence of two colonizations, one into the southern Lesser Antilles and the second producing the diverse adaptive radiations throughout the rest of the Caribbean. Within this latter clade, anoles initially diversified primarily on the two larger Greater Antillean islands (although Puerto Rico also seems to have been involved) before subsequently undergoing secondary radiations on all of the islands and eventually returning to the mainland, where this back-colonization has produced an extensive evolutionary radiation. The phylogeny also indicates that very few inter-island dispersal events occurred in Greater Antillean evolution. Rather, the Greater Antillean faunas, renowned for the extent to which the same ecomorphs are found on each island, are primarily the result of convergent evolution²⁵.

The genome sequence of *A. carolinensis* allows a deeper understanding of amniote evolution. Filling this important reptilian node with a sequenced genome has revealed derived states in each major amniote branch and has helped to illuminate the amniote ancestor. However, the tree of sequenced reptilian genomes is still extremely sparse, and the sequencing of additional non-avian reptiles would be necessary to fully understand how typical *A. carolinensis* and the sequenced bird genomes are of the entire reptile clade.

In addition to the utility of the *A. carolinensis* genome sequence as a representative of non-avian reptiles, *Anolis* species are a unique resource for the study of adaptive radiation and convergent evolution. With their invasions of and subsequent radiations on Caribbean islands, anoles provide a terrestrial analogue to stickleback and cichlid fish, which underwent adaptive evolution in separate aquatic environments. Just as genomic research in sticklebacks has deepened the study of aquatic ecological speciation, a large-scale genomic phylogenetic survey of the Caribbean anoles would be an opportunity for detailed study of adaptive evolution in a land animal²⁶; in particular because anole genomes contain large numbers of active mobile elements that we speculate could form substrates for exaptation of novel regulatory elements.

METHODS SUMMARY

A full description of methods, including sample collection, sequencing, assembly, anchoring, mass spectrometry and all sequence analysis, can be found in Supplementary Information. All animal experiments were approved by the MIT Committee for Animal Care.

Received 21 April; accepted 27 July 2011.

Published online 31 August 2011.

- Sumida, S. S. & Martin, K. L. M. *Amniote Origins: Completing the Transition to Land* (Academic Press, 1997).
- Hillier, L. W. *et al.* Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695–716 (2004).
- Warren, W. C. *et al.* The genome of a songbird. *Nature* **464**, 757–762 (2010).
- Dalloul, R. A. *et al.* Multi-platform next-generation sequencing of the domestic turkey (*Meleagris gallopavo*): genome assembly and analysis. *PLoS Biol.* **8**, e1000475 (2010).
- Fujita, M. K., Edwards, S. V. & Ponting, C. P. Unexpected homogeneity in GC content and divergent isochore structure in the *Anolis* genome. *Genome Biol. Evol.* doi:10.1093/gbe/evr072 (in the press).
- Benton, M. J. & Donoghue, P. C. Paleontological evidence to date the tree of life. *Mol. Biol. Evol.* **24**, 26–53 (2007).
- Axelsson, E., Webster, M. T., Smith, N. G. C., Burt, D. W. & Ellegren, H. Comparison of the chicken and turkey genomes reveals a higher rate of nucleotide divergence on microchromosomes than macrochromosomes. *Genome Res.* **15**, 120–125 (2005).

- Burt, D. W. Origin and evolution of avian microchromosomes. *Cytogenet. Genome Res.* **96**, 97–112 (2002).
- Potrzebowski, L. *et al.* Chromosomal gene movements reflect the recent origin and biology of therian sex chromosomes. *PLoS Biol.* **6**, e80 (2008).
- Organ, C. L. & Janes, D. E. Evolution of sex chromosomes in Saurapsida. *Integr. Comp. Biol.* **48**, 512–519 (2008).
- Losos, J. B. *Lizards in an Evolutionary Tree: Ecology and Adaptive Radiation of Anoles* (Univ. of California Press, 2009).
- Gorman, G. C. in *Cytotaxonomy and Vertebrate Evolution* (eds Chiarelli, A. B. & Capanna, E.) Ch. 12 349–424 (Academic, 1973).
- Hellsten, U. *et al.* The genome of the Western clawed frog *Xenopus tropicalis*. *Science* **328**, 633–636 (2010).
- Bininda-Emonds, O. R. *et al.* The delayed rise of present-day mammals. *Nature* **446**, 507–512 (2007).
- Shedlock, A. M. *et al.* Phylogenomics of nonavian reptiles and the structure of the ancestral amniote genome. *Proc. Natl Acad. Sci. USA* **104**, 2767–2772 (2007).
- Kordis, D. Transposable elements in reptilian and avian (sauropsida) genomes. *Cytogenet. Genome Res.* **127**, 94–111 (2009).
- Novick, P. A., Basta, H., Floumanhaft, M., McClure, M. A. & Boissinot, S. The evolutionary dynamics of autonomous non-LTR retrotransposons in the lizard *Anolis carolinensis* shows more similarity to fish than mammals. *Mol. Biol. Evol.* **26**, 1811–1822 (2009).
- Novick, P. A., Smith, J. D., Floumanhaft, M., Ray, D. A. & Boissinot, S. The evolution and diversity of DNA transposons in the genome of the lizard *Anolis carolinensis*. *Genome Biol. Evol.* **3**, 1–14 (2010).
- Furano, A. V., Duvernell, D. D. & Boissinot, S. L1 (LINE-1) retrotransposon diversity differs dramatically between mammals and fish. *Trends Genet.* **20**, 9–14 (2004).
- Duret, L., Eyre-Walker, A. & Galtier, N. A new perspective on isochore evolution. *Gene* **385**, 71–74 (2006).
- Viets, B. E., Ewert, M. A., Talent, L. G. & Nelson, C. E. Sex-determining mechanisms in squamate reptiles. *J. Exp. Zool.* **270**, 45–56 (1994).
- Mann, K. The chicken egg white proteome. *Proteomics* **7**, 3558–3568 (2007).
- Farinazzo, A. *et al.* Chicken egg yolk cytoplasmic proteome, mined via combinatorial peptide ligand libraries. *J. Chromatogr. A* **1216**, 1241–1252 (2009).
- Lowe, C. B., Bejerano, G., Salama, S. R. & Haussler, D. Endangered species hold clues to human evolution. *J. Hered.* **101**, 437–447 (2010).
- Williams, E. E. in *Lizard Ecology: Studies of a Model Organism* (eds Huey, R. B., Pianka, E. R. & Schoener, T. W.) (Harvard Univ. Press, 1983).
- Colosimo, P. F. *et al.* Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science* **307**, 1928–1933 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Generation of the *A. carolinensis* sequence at the Broad Institute of MIT and Harvard was supported by grants from the National Human Genome Research Institute (NHGRI); S.P. and R.E.G. were supported by NSF grants DEB-0844624 and DEB-0920892. All sequence data was produced by the Genome Sequencing Platform of the Broad Institute. We would also like to thank the David and Lucile Packard Foundation for their early support of anole genomics, R. Andrews for her advice on lizard egg biology, C. Hickman and B. Temple and the Herpetology group at the University of Georgia's Savannah River Ecology Lab for assistance with sample collection, and L. Gaffney and L. Virnoche for assistance with figure and text preparation.

Author Contributions J.A., F.D. and K.L.-T. planned and oversaw the project. M.G., D. Heiman and S.Y. assembled the genome. B.F.T.H., M.Y.K. and P.J.d.J. constructed the BAC library. T.C.G. and J.W. provided tissues for sequencing libraries and FISH analysis. M.B., C.W. and D. Heiman anchored the genome. T.A.C. and D.D.P. assembled the mitochondrial genome. J.K.C. and Z.S. constructed the cDNA library. S.S. and A.Z. annotated the genome. L.K., A.H. and C.P.P. performed the gene repertoire analysis. T.S. aided egg protein experimental design. J.D.J. and S.E.P. performed egg mass spectrometry. M.G. performed genome synteny analysis. E.M. performed segmental duplication analysis. C.W. and M.B. discovered the sex chromosomes and the pericentromeric inversions. P.R. performed the microchromosome and GC analysis. M.K.F. and C.P.P. participated in microchromosome and GC data interpretation. D.A.R. constructed the repeat library. D.A.R., S.B., P.A.N., A.M.S., J.D.S. and C.B. performed the repeat analysis. M.G., J.B.L., R.E.G., S.P., K.d.Q. and R.S. participated in phylogeny data collection. R.E.G., S.P., K.d.Q. and R.S. participated in phylogeny analysis. All authors participated in data discussion and interpretation. J.A., F.D., C.B.L., R.G., D.A.R., S.V.E., C.J.S., J.B.L., E.S.L., M.B., C.P.P. and K.L.-T. wrote the paper with input from other authors.

Author Information The *A. carolinensis* whole-genome shotgun project has been deposited in GenBank under the project accession AAWZ00000000.2. All phylogeny sequence data can be found at <http://purl.org/phylo/treebase/phyloids/study/TB2:S11713>. Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and is freely available to all readers at www.nature.com/nature. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.A. (jalfoldi@broadinstitute.org) or K.L.-T. (kersli@broadinstitute.org).

Innate immune recognition of bacterial ligands by NAIPs determines inflammasome specificity

Eric M. Kofoed¹ & Russell E. Vance¹

Inflammasomes are a family of cytosolic multiprotein complexes that initiate innate immune responses to pathogenic microbes by activating the caspase 1 protease^{1,2}. Although genetic data support a critical role for inflammasomes in immune defence and inflammatory diseases³, the molecular basis by which individual inflammasomes respond to specific stimuli remains poorly understood. The inflammasome that contains the NLRC4 (NLR family, CARD domain containing 4) protein was previously shown to be activated in response to two distinct bacterial proteins, flagellin^{4,5} and PrgJ⁶, a conserved component of pathogen-associated type III secretion systems. However, direct binding between NLRC4 and flagellin or PrgJ has never been demonstrated. A homologue of NLRC4, NAIP5 (NLR family, apoptosis inhibitory protein 5), has been implicated in activation of NLRC4 (refs 7–11), but is widely assumed to have only an auxiliary role^{1,2}, as NAIP5 is often dispensable for NLRC4 activation^{7,8}. However, *Naip5* is a member of a small multigene family¹², raising the possibility of redundancy and functional specialization among *Naip* genes. Here we show in mice that different NAIP paralogues determine the specificity of the NLRC4 inflammasome for distinct bacterial ligands. In particular, we found that activation of endogenous NLRC4 by bacterial PrgJ requires NAIP2, a previously uncharacterized member of the NAIP gene family, whereas NAIP5 and NAIP6 activate NLRC4 specifically in response to bacterial flagellin. We dissected the biochemical mechanism underlying the requirement for NAIP proteins by use of a reconstituted NLRC4 inflammasome system. We found that NAIP proteins control ligand-dependent oligomerization of NLRC4 and that the NAIP2–NLRC4 complex physically associates with PrgJ but not flagellin, whereas NAIP5–NLRC4 associates with flagellin but not PrgJ. Our results identify NAIPs as immune sensor proteins and provide biochemical evidence for a simple receptor–ligand model for activation of the NAIP–NLRC4 inflammasomes.

A fundamental question in immunology is how host defence is initiated in response to specific microbial ligands. The inflammasome containing the NLRC4 protein activates caspase 1 (CASP1) in response to the carboxy terminus of bacterial flagellin^{6,7}, as well as in response to the inner rod protein of the type III secretion systems of diverse bacterial species (for example, PrgJ of *Salmonella enterica* serovar Typhimurium)⁶. Activated CASP1 processes interleukin (IL)-1 β and IL-18 inflammatory cytokines and induces a rapid and inflammatory host cell death called pyroptosis¹³. In certain cases, NLRC4 activation requires NAIP5, as *Naip5*^{−/−} mice fail to activate NLRC4 or CASP1 in response to infection with *Legionella pneumophila* or in response to the C terminus of flagellin^{7,8}. Interestingly, however, NAIP5 is not essential for NLRC4 activation in response to *S. enterica* Typhimurium or PrgJ^{7,8}.

In addition to *Naip5*, C57BL/6 mice express three other *Naip* genes (*Naip1*, *Naip2* and *Naip6*), the functions of which remain unknown¹². We hypothesized that each NAIP paralogue may have evolved to be specific for a unique bacterial ligand. We first focused on NAIP2, as it appeared to be highly expressed in C57BL/6 mice¹⁴. We used specific

short hairpin RNAs (shRNAs) to knock down *Naip2* expression in primary bone-marrow-derived macrophages. ShRNA1 and shRNA2 specifically reduced NAIP2 protein levels without targeting other NAIP paralogues, whereas empty vector, shRNA3 or a scrambled control shRNA had little effect on NAIP2 protein levels (Supplementary Fig. 1a, b). Macrophages expressing these shRNAs were then infected with flagellin-deficient *Listeria* strains that inducibly express PrgJ (*Listeria*-PrgJ) or flagellin (*Listeria*-FlaA)⁸. A *Listeria*-based system was chosen because it is an efficient means for delivering PrgJ to macrophages⁸, and because it allows for controlled comparisons of PrgJ and FlaA within a single experimental system. Notably, knockdown of *Naip2* prevented pyroptosis and CASP1 activation by *Listeria*-PrgJ (Fig. 1a–c). By contrast, *Naip2* knockdown did not affect inflammasome activation by *Listeria*-FlaA (Fig. 1b, c) or *L. pneumophila*, which expresses flagellin but not PrgJ (Supplementary Fig. 1c). Instead, flagellin-dependent inflammasome activation depended on *Naip5*, as previously shown^{7–11}. Inflammasome activation by wild-type *Salmonella*, which encodes both flagellin and PrgJ, was not significantly affected by *Naip2* knockdown (Fig. 1d, e). However, knockdown of *Naip2* in *Naip5*^{−/−} macrophages significantly reduced or abolished inflammasome activation by wild-type *Salmonella* (Fig. 1d, e), indicating that both NAIP2 and NAIP5 recognize *Salmonella*. Interestingly, inflammasome activation by flagellin-deficient (FliC[−] FljB[−]) *Salmonella*, which still express PrgJ, depended entirely on *Naip2* (Fig. 1d, e). Taken together, these data indicate that *Naip2* is specifically required for activation of the NLRC4 inflammasome by PrgJ, in contrast to *Naip5*, which seems to be specifically required for NLRC4 activation by flagellin.

Biochemical analysis of the inflammasome in macrophages is complicated by the expression of multiple NAIP proteins and by their low expression levels. We therefore decided to reconstitute the NLRC4 inflammasome in non-immune 293T cells, which do not express NLRC4 or NAIPs, so that the functions of individual NAIP proteins could be analysed. 293T cells transiently transfected with green fluorescent protein (GFP)-marked vectors encoding wild-type NLRC4, NAIP5 and CASP1 did not exhibit significant spontaneous inflammasome activation, and instead, most cells expressed GFP (Fig. 2a). However, when flagellin (FlaA) from *L. pneumophila* was co-expressed with NLRC4, NAIP5 and CASP1, we observed a significant loss of GFP^{high} cells and an increase in the number of dead (7AAD⁺) cells (Fig. 2a). This result was highly reminiscent of flagellin-dependent activation of the endogenous NAIP5–NLRC4 inflammasome in macrophages, which also results in a rapid CASP1-dependent cell death, loss of membrane integrity, and release of cytosolic contents and GFP⁷. Similar to the genetic requirement for *Nlrc4*, *Naip5* and *Casp1* in macrophages^{4,5,7–11,15}, we found that NAIP5, NLRC4, catalytically active CASP1, and FlaA are all required to trigger cell death and loss of membrane integrity/GFP in reconstituted 293T cells (Fig. 2b, c). The reconstituted NAIP5–NLRC4 inflammasome also recapitulated the ability of native inflammasomes to process CASP1 and IL-1 β in response to cytosolic flagellin (Supplementary Fig. 2). Consistent with a lack of a role for NAIP5 in recognition of PrgJ by macrophages⁸, the reconstituted NAIP5–NLRC4 inflammasome did not respond to PrgJ (Fig. 2d, e). By

¹Division of Immunology and Pathogenesis, Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA.

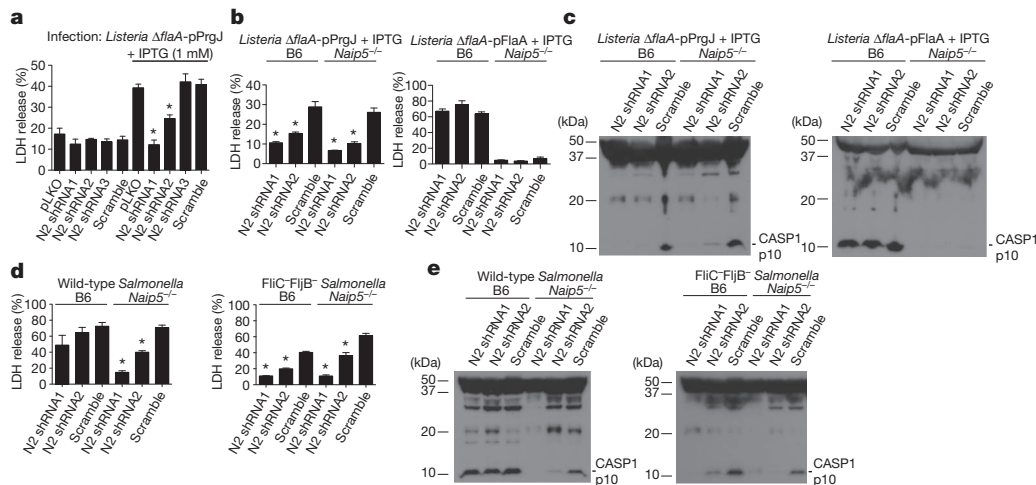


Figure 1 | NAIP2 is required in macrophages for inflammasome activation in response to PrgJ. **a–c**, Primary bone-marrow-derived macrophages expressing shRNAs targeting *Naip2* (N2) (or controls) were infected with flagellin-deficient *L. monocytogenes* (multiplicity of infection = 5) expressing a secreted ActA100–PrgJ (pPrgJ) or ActA100–FlaA (pFlaA) fusion protein under IPTG-inducible control. Cell death (\pm s.d.) was measured in triplicate by LDH release 6 h after infection (**a**, **b**), or active CASP1 (p10) was measured by

western blotting of cell supernatants (**c**). **d**, **e**, NAIP2 knockdown cells were infected with wild-type or flagellin-deficient (FlaA⁻ *S. enterica* Typhimurium) and inflammasome activation was measured by LDH release (\pm s.d.) at 3 h after infection (**d**) or CASP1 processing (**e**). Data shown are representative of two (**c**, **e**) or three (**a**, **b**, **d**) independent experiments. **P* < 0.02 as compared to scramble (Student's *t*-test, two-tailed).

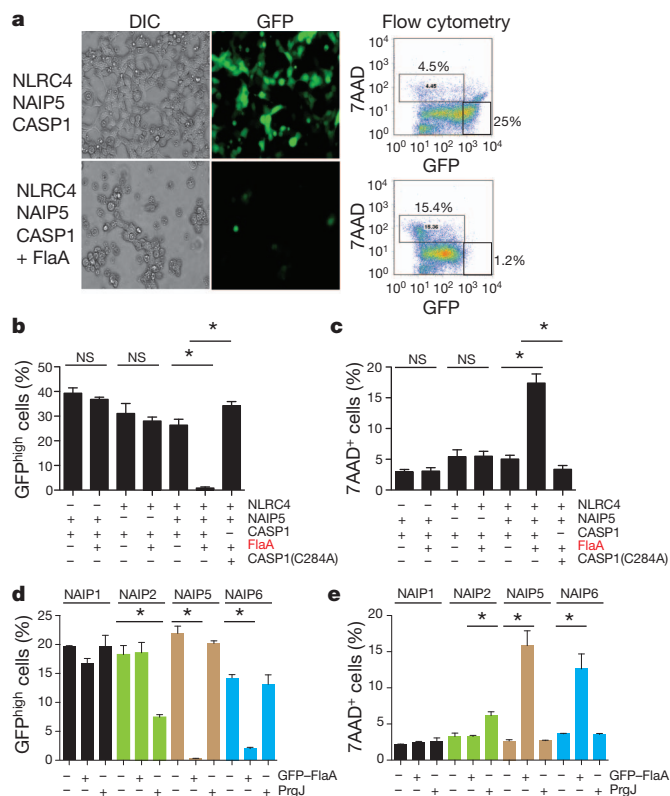


Figure 2 | Reconstitution of the NAIP5–NLRC4 inflammasome in 293T cells. **a**, GFP-marked expression vectors encoding NLRC4, NAIP5, CASP1 and/or flagellin (FlaA) were transiently transfected into 293T cells. Cells were imaged for differential interference contrast (DIC) and GFP fluorescence 48 h later. Dead cells were stained with 7AAD. **b**, **c**, GFP^{high} cells (**b**) and 7AAD^{high} cells (**c**) were quantified (\pm s.d.) as in **a**, but with specific expression vectors omitted from the transfection as indicated. CASP1(C284A) is a catalytically dead mutant. **d**, **e**, 293T cells were transfected as indicated and analysed as above. Data shown (\pm s.d.) are representative of at least three independent experiments. **P* < 0.02 (Student's *t*-test, two tailed); NS, not significant.

contrast, a reconstituted NAIP2–NLRC4 inflammasome responded specifically to PrgJ but not flagellin (Fig. 2d, e). Taken together, we conclude that we have successfully reconstituted NAIP2–NLRC4 and NAIP5–NLRC4 inflammasomes that exhibit all the known requirements and specificities of the native inflammasomes.

It is believed that activated inflammasomes assemble into high-molecular-mass multiprotein complexes¹⁶, but this has not been demonstrated for the NLRC4 inflammasome. To visualize inflammasome assembly, 293T cells were transfected with NAIP5, NLRC4 and FlaA in various combinations, but CASP1 was omitted so that cell death and loss of cellular contents (and assembled inflammasomes) would not occur. Digitonin-solubilized cell lysates were resolved on blue native (BN) polyacrylamide gel electrophoresis (PAGE) gels¹⁷. A marked shift of NLRC4 from a monomer (~120 kDa) to an oligomeric complex (~1,000 kDa) was seen in the presence of NAIP5 and FlaA. NAIP5 was also contained within the high-molecular-mass oligomeric complex (Fig. 3a). The association of NAIP5 and NLRC4 in the same complex was validated by co-immunoprecipitation (Supplementary Fig. 3)^{11,18}. NLRC4 oligomerization was induced by either untagged FlaA or a GFP–FlaA fusion protein (Fig. 3a), both of which activate NLRC4–CASP1. Importantly, assembly of the NLRC4 inflammasome required FlaA (Supplementary Fig. 4a) and was not observed in the absence of NAIP5 (Fig. 3a), indicating that a biochemical function of NAIP5 is to promote NLRC4 oligomerization.

Despite strong genetic evidence that NLR proteins, such as NAIP5 and NLRC4, function as microbial ‘sensors’, there is no biochemical evidence that NLRs interact directly with microbial ligands. In fact, some studies of the NLRP3 inflammasome^{19–21}, as well as analyses of analogous proteins from plants²², suggest that at least some NLRs recognize pathogens indirectly. To determine if the oligomerized NAIP5–NLRC4 complex also contains flagellin, we subjected samples separated in the first dimension by native PAGE to a second dimension of SDS–PAGE. To facilitate detection of flagellin, we used a 6×Myc-tagged flagellin, which activates the inflammasome identically to native flagellin (data not shown). This approach revealed that FlaA was indeed present in a high-molecular-mass complex, along with NAIP5 and NLRC4 (Fig. 3b). NAIP5 exhibited a weak flagellin-dependent mobility shift in the absence of NLRC4 (Supplementary Fig. 4b), indicating that NLRC4 is not essential for flagellin recognition, although formation/stabilization of the oligomerized complex seems to be significantly enhanced by NLRC4. FlaA expressed alone was present in cell extracts

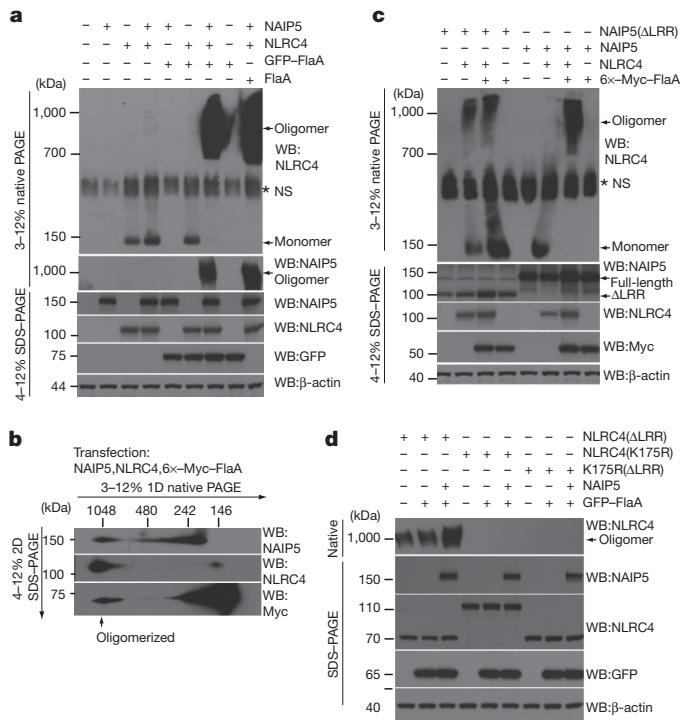


Figure 3 | NAIP5 is required for formation of a hetero-oligomeric complex that contains NLRC4, NAIP5 and flagellin. **a**, 293T cells were transfected as indicated, followed by analysis by blue native PAGE or SDS-PAGE, and western blotting. *NS, nonspecific band. **b**, 293T cells were transfected as indicated and lysates were separated by a first dimension of blue native PAGE followed by a second dimension of SDS-PAGE. **c**, **d**, 293T cells were transfected as indicated and samples were processed and analysed as in **a**. Data shown are representative of at least three independent experiments.

only as a monomer (Supplementary Fig. 4c). Taken together, these observations provide evidence for a simple receptor–ligand model of NAIP5–NLRC4 activation by flagellin.

Consistent with the autoinhibitory function of the leucine-rich repeats (LRRs) in other NLRs, we found that NAIP5(ΔLRR) and NLRC4(ΔLRR) constitutively activated CASP1-dependent cell death, independent of the presence of flagellin (Supplementary Fig. 5). Interestingly, NLRC4(ΔLRR) was able to activate CASP1 in the absence of NAIP5, whereas constitutively active NAIP5(ΔLRR) required wild-type NLRC4 to activate CASP1. This result suggests that NAIP5 functions upstream of NLRC4. Indeed, NAIP5(ΔLRR) was able to induce the oligomerization of wild-type NLRC4 (Fig. 3c), whereas the spontaneous oligomerization of NLRC4(ΔLRR) did not require NAIP5 (Fig. 3d). Spontaneous oligomerization of NLRC4(ΔLRR) did require the nucleotide-binding domain (NBD) of NLRC4, as a K175R mutation previously shown to disrupt NBD function²³ abolished NLRC4(ΔLRR) auto-oligomerization (Fig. 3d). The ability of NAIP5 to induce oligomerization of NLRC4 in response to flagellin required both the NBD and amino-terminal BIRs of NAIP5, but did not require the N-terminal CARD of NLRC4 (Supplementary Fig. 4d, e), whereas functional CASP1 activation required all of these domains (Supplementary Figs 5 and 6). Taken together, these data are indicative of a working model (Supplementary Fig. 7) in which NAIP5 is activated by flagellin and induces downstream NLRC4 oligomerization and CASP1 activation.

Consistent with a specific role for NAIP2 in recognition of PrgJ, we found that PrgJ did not induce the oligomerization of NAIP5–NLRC4 (Fig. 4a) but did induce oligomerization of NAIP2–NLRC4. Oligomerization of NLRC4 did not occur when co-expressed with NAIP2 alone or with NAIP2 and FlaA (Fig. 4b). Interestingly, NAIP6 resembled NAIP5 and supported NLRC4 oligomerization in response to FlaA but

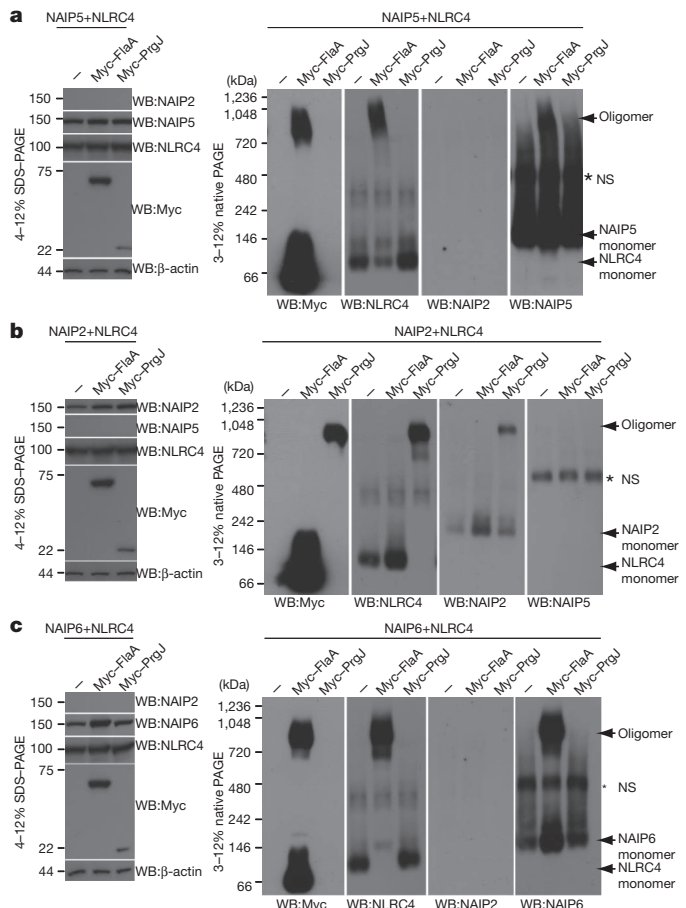


Figure 4 | NAIP paralogues confer specificity to the NLRC4 inflammasome.

a, 293T cells were co-transfected with wild-type NAIP5 and NLRC4, alone or in combination with 6x-Myc-FlaA or 6x-Myc-PrgJ followed by blue native PAGE 48 h later. *NS, nonspecific band. Whole-cell lysates were also separated by conventional 4–12% SDS-PAGE to control for expression of each transfected gene construct (left panel). **b**, 293T cells were transfected with wild-type NAIP2 and NLRC4 and analysed as in **a**. **c**, 293T cells were transfected with wild-type NAIP6 and NLRC4, and analysed as in **a**. Data shown are representative of at least three independent experiments.

not PrgJ (Fig. 4c), perhaps providing an explanation for the previously puzzling observation that *Naip5*^{−/−} cells can respond to high levels of flagellin⁷. In contrast, NAIP1 is an ‘orphan’ NAIP because it responded neither to PrgJ nor flagellin (Fig. 2d, e and Supplementary Fig. 8).

Our results demonstrate that the ability of the NLRC4 inflammasome to assemble and functionally activate CASP1 in response to specific bacterial ligands is dictated by NAIP family members. The most parsimonious model to account for our results is that NAIP proteins function as direct receptors for bacterial ligands (Supplementary Fig. 7). Although NLRC4 was previously suspected to be the cytosolic flagellin sensor^{1,2}, we hypothesize that a main function of NLRC4 may instead be to serve as an adaptor, downstream of NAIP proteins, to recruit CASP1 via a CARD–CARD interaction. NLRC4 may also have an important function in ligand binding or in stabilizing NAIP–NLRC4–ligand complexes, but the specificity of the complexes for particular ligands seems to be controlled by NAIP proteins.

The number and sequence of *Naip* paralogues vary significantly among inbred mouse strains, and have been suggested to be evolving rapidly²⁴. Indeed, the murine *Naip* locus was originally identified by a forward genetic approach which took advantage of the widely varying susceptibility of inbred mouse strains to *L. pneumophila* infection^{14,25}. The single known human NAIP orthologue may also exist within a rapidly evolving locus²⁴; our results indicate that it will be of great interest to establish the specificity of the human NAIP protein. We

propose that *Naip* gene evolution represents a fascinating example of the molecular arms race between bacteria and their hosts.

METHODS SUMMARY

Naip2 knockdown. Primary C57BL/6 bone marrow cells were transduced with pLKO.1-based lentivirus encoding shRNAs that specifically target *Naip2* or controls. Bone marrow cells were differentiated into macrophages by culture in media containing macrophage colony-stimulating factor (M-CSF). On day 4 of culture, transduced macrophages were selected by addition of puromycin ($5 \mu\text{g ml}^{-1}$). On day 8 of culture, macrophages were re-plated and infected the next day with *Listeria monocytogenes* or *S. enterica* Typhimurium expressing flagellin or PrgJ⁸, and inflammasome activation was measured by assaying release of the cytosolic enzyme lactate dehydrogenase (LDH)⁷ or by western blotting for processed (p10) CASP1.

Reconstituted inflammasome. The inflammasome was reconstituted by transfection of 293T cells with MSCV2.2-IRES-GFP-based expression vectors encoding various mouse (C57BL/6-derived) *Naip* genes, *Nlr4* and caspase 1. Inflammasome oligomerization was assessed in digitonin (1%) lysates using a Bis-Tris NativePAGE system (Invitrogen) followed by western blotting.

Statistical analysis. Statistical differences were calculated with an unpaired two-tailed Student's *t*-test using GraphPad Prism 5.0b.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 27 June; accepted 27 July 2011.

Published online 28 August 2011.

- Latz, E. The inflammasomes: mechanisms of activation and function. *Curr. Opin. Immunol.* **22**, 28–33 (2010).
- Schroder, K. & Tschopp, J. The inflammasomes. *Cell* **140**, 821–832 (2010).
- Ting, J. P., Kastner, D. L. & Hoffman, H. M. CATERPILLERS, pyrin and hereditary immunological disorders. *Nature Rev. Immunol.* **6**, 183–195 (2006).
- Franchi, L. *et al.* Cytosolic flagellin requires Ipaf for activation of caspase-1 and interleukin 1 β in salmonella-infected macrophages. *Nature Immunol.* **7**, 576–582 (2006).
- Miao, E. A. *et al.* Cytoplasmic flagellin activates caspase-1 and secretion of interleukin 1 β via Ipaf. *Nature Immunol.* **7**, 569–575 (2006).
- Miao, E. A. *et al.* Innate immune detection of the type III secretion apparatus through the NLRC4 inflammasome. *Proc. Natl Acad. Sci. USA* **107**, 3076–3080 (2010).
- Lightfield, K. L. *et al.* Critical function for Naip5 in inflammasome activation by a conserved carboxy-terminal domain of flagellin. *Nature Immunol.* **9**, 1171–1178 (2008).
- Lightfield, K. L. *et al.* Differential requirements for NAIP5 in activation of the NLRC4 (IPAF) inflammasome. *Infect. Immun.* **79**, 1606–1614 (2011).
- Molofsky, A. B. *et al.* Cytosolic recognition of flagellin by mouse macrophages restricts *Legionella pneumophila* infection. *J. Exp. Med.* **203**, 1093–1104 (2006).
- Ren, T., Zamboni, D. S., Roy, C. R., Dietrich, W. F. & Vance, R. E. Flagellin-deficient *Legionella* mutants evade caspase-1- and Naip5-mediated macrophage immunity. *PLoS Pathog.* **2**, e18 (2006).
- Zamboni, D. S. *et al.* The Birc1e cytosolic pattern-recognition receptor contributes to the detection and control of *Legionella pneumophila* infection. *Nature Immunol.* **7**, 318–325 (2006).
- Growney, J. D. & Dietrich, W. F. High-resolution genetic and physical map of the Lgn1 interval in C57BL/6J implicates Naip2 or Naip5 in *Legionella pneumophila* pathogenesis. *Genome Res.* **10**, 1158–1171 (2000).
- Bergsbaken, T., Fink, S. L. & Cookson, B. T. Pyroptosis: host cell death and inflammation. *Nature Rev. Microbiol.* **7**, 99–109 (2009).
- Wright, E. K. *et al.* Naip5 affects host susceptibility to the intracellular pathogen *Legionella pneumophila*. *Curr. Biol.* **13**, 27–36 (2003).
- Broz, P., von Moltke, J., Jones, J. W., Vance, R. E. & Monack, D. M. Differential requirement for Caspase-1 autoproteolysis in pathogen-induced cell death and cytokine processing. *Cell Host Microbe* **8**, 471–483 (2010).
- Martinon, F., Burns, K. & Tschopp, J. The inflammasome: a molecular platform triggering activation of inflammatory caspases and processing of proIL- β . *Mol. Cell* **10**, 417–426 (2002).
- Schagger, H., Cramer, W. A. & von Jagow, G. Analysis of molecular masses and oligomeric states of protein complexes by blue native electrophoresis and isolation of membrane protein complexes by two-dimensional native electrophoresis. *Anal. Biochem.* **217**, 220–230 (1994).
- Damiano, J. S., Oliveira, V., Welsh, K. & Reed, J. C. Heterotypic interactions among NACHT domains: implications for regulation of innate immune responses. *Biochem. J.* **381**, 213–219 (2004).
- Hornung, V. *et al.* Silica crystals and aluminum salts activate the NALP3 inflammasome through phagosomal destabilization. *Nature Immunol.* **9**, 847–856 (2008).
- Zhou, R., Tardivel, A., Thorens, B., Choi, I. & Tschopp, J. Thioredoxin-interacting protein links oxidative stress to inflammasome activation. *Nature Immunol.* **11**, 136–140 (2010).
- Zhou, R., Yazdi, A. S., Menu, P. & Tschopp, J. A role for mitochondria in NLRP3 inflammasome activation. *Nature* **469**, 221–225 (2011).
- Chisholm, S. T., Coaker, G., Day, B. & Staskawicz, B. J. Host-microbe interactions: shaping the evolution of the plant immune response. *Cell* **124**, 803–814 (2006).
- Lu, C. *et al.* Nucleotide binding to CARD12 and its role in CARD12-mediated caspase-1 activation. *Biochem. Biophys. Res. Commun.* **331**, 1114–1119 (2005).
- Romanish, M. T., Lock, W. M., de Lagaemaat, L. N., Dunn, C. A. & Mager, D. L. Repeated recruitment of LTR retrotransposons as promoters by the anti-apoptotic locus *NAIP* during mammalian evolution. *PLoS Genet.* **3**, e10 (2007).
- Diez, E. *et al.* *Birc1e* is the gene within the *Lgn1* locus associated with resistance to *Legionella pneumophila*. *Nature Genet.* **33**, 55–60 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Work in R.E.V.'s laboratory is supported by Investigator Awards from the Burroughs Wellcome Fund and the Cancer Research Institute and by NIH grants AI075039, AI080749 and AI063302. We thank J. von Moltke, A. Kintzer and B. Krantz for provision of reagents; S. Mariathasan and V. Dixit for the gift of anti-NLRC4 antibodies and *Nlr4*^{-/-} mice; J. D. Sauer and D. Portnoy for development of *Listeria* strains to deliver PrgJ and FlaA; and E. Michelle Long and W. Dietrich for pCDNA3-NAIP constructs and anti-NAIP antibodies. We thank A. Roberts for her initial efforts to knock down NAIP2, M. Fontana for validating NAIP knockdowns, and J. von Moltke and members of the Barton and Vance laboratories for discussions.

Author Contributions E.M.K. and R.E.V. conceived the experiments and wrote the paper. E.M.K. performed the experiments.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.E.V. (rvance@berkeley.edu).

METHODS

Mice. C57BL/6J (B6) mice were purchased from Jackson Labs and bred at UC Berkeley. *Naip5*^{-/-} mice on a pure B6 background were described previously⁷. Animal experiments were approved by the UC Berkeley Animal Care and Use Committee.

Cell culture. HEK293T cells were grown in complete media (DMEM, 10% FBS, 100 units ml⁻¹ penicillin, 100 µg ml⁻¹ streptomycin, 2 mM L-glutamine). Bone marrow macrophages were differentiated from bone-marrow-derived precursor cells using macrophage colony stimulating factor as described previously⁷.

Transient transfections. HEK293T cells were plated at 8 × 10⁵ cells per well in 6-well plates, transfected the next day using Lipofectamine 2000, and collected for flow cytometric analysis 48 h later.

Measurement of cell death. Cell death of HEK293T cells was measured by flow cytometry measuring GFP and 7AAD fluorescence. Cells were stained with 7AAD (BD Pharmingen) according to the manufacturer's instructions. Death of immortalized macrophages in response to *L. pneumophila* and LFn-FlaA was measured by following release of the intracellular enzyme lactate dehydrogenase (LDH) as described previously⁷. Infection of immortalized macrophages with *L. pneumophila* was performed using a multiplicity of infection (MOI) of 2, and cell death was measured 4 h later as described previously⁷. LFn-FlaA is a recombinant 6×His-tagged fusion protein encoding the first (non-enzymatically active) 263 amino acids of lethal factor from *Bacillus anthracis* fused to full-length flagellin (FlaA) from *L. pneumophila* (J. von Moltke and R.E.V., unpublished). LFn-FlaA and *B. anthracis* protective antigen (PA) were purified from *E. coli* as described previously²⁶. Endotoxin was removed from these proteins using Detoxi-Gel (Pierce) according to the manufacturer's protocol.

Western blotting and native PAGE. Blue native gel electrophoresis was performed using the Bis-Tris NativePAGE system by Invitrogen according to the manufacturer's instructions. Briefly, HEK293T cells were transfected for 48 h, washed twice with cold PBS, trypsinized for 3 min at 37 °C, re-suspended in complete media, and pelleted by centrifugation at 400g for 5 min at 4 °C. Cell pellets were washed twice with cold PBS, followed by re-suspension in 1% digitonin native lysis buffer (50 mM BisTris, 50 mM NaCl, 10% w/v glycerol, 0.001% Ponceau S, 1% digitonin, 2 mM Na₂VO₄, 1 mM PMSF, 25 mM NaF, 1× Roche protease inhibitor cocktail (no EDTA), pH 7.2). Cell lysates were triturated and incubated on ice for 30 min, and then insoluble cell debris was pelleted by centrifugation at 16,000g for 30 min at 4 °C. Lysates were quantified for total protein using the BCA protein assay (Pierce), equalized for total protein, and separated by NativePAGE using the Novex BisTris gel system according to the manufacturer's instructions (Invitrogen). Native gels were soaked in 10% SDS for 5 min before transfer to PVDF membrane (Millipore) and conventional western blotting. Antibodies used were anti-mIL-1β (R&D systems), anti-caspase 1 p10 (M20) (Santa Cruz), anti-NLRC4 (gift of S. Mariathasan and V. Dixit, Genentech), anti-NAIP5(961–978)¹⁴, anti-Myc (9E10) (Clontech), anti-β-actin (C4) (Santa Cruz), anti-GFP (JL-8) (Clontech), anti-NAIP2(33–46)¹⁴, anti-rabbit IgG-HRP and anti-mouse IgG-HRP (GE Healthcare), anti-goat IgG-HRP (Santa Cruz). In some cases, native gels were subjected to a second dimension of electrophoresis using SDS–PAGE. A 5.7-cm slice (lane) of natively resolved gel was placed in a dish containing 1× Laemmli sample buffer (50 mM Tris-Cl (pH 6.8), 100 mM DTT, 2% (w/v) SDS, 0.1% bromophenol blue, 10% (v/v) glycerol) for 10 min, microwaved on high for 20 s, and rocked for another 5 min before loading slice into the well of a precast 2D 4–12% SDS–PAGE gel (Invitrogen). For immunoprecipitations, digitonin cell extracts (100 µg total protein) were pre-cleared with 25 µl of washed Protein-G-sepharose (GE Healthcare), and then cleared extracts were incubated with 1 µg primary antibody (or isotype controls) overnight at 4 °C, and complexes were captured the following day with 25 µl of washed Protein-G-sepharose. Bound proteins were eluted by re-suspension in Laemmli sample buffer, boiled for 5 min and separated by SDS–PAGE.

Expression constructs. NAIP5 wild-type and mutant constructs were cloned into the MSCV2.2 retroviral vector containing an IRES–GFP downstream of the multiple cloning site. Expression is driven from the viral LTR and is considerably lower than that driven by the CMV promoter (data not shown). In general, wild-type and mutant ORFs were amplified between flanking BamHI and NotI sites, and a Kozak sequence (GCCACC) was engineered to precede the start codon. The BamHI/NotI digested PCR product was cloned into complementary BglII/NotI digested MSCV2.2 vector. Wild-type NAIP5(1–1402) and NAIP5(Δloop) (Δ464–487) were amplified from pcDNA3 using forward (AAAAGGATCCGCCACCATGGC TGAGCATGGGGAGTCTCCG) and reverse (AAAAGCGCCGCTTACTCC AGGATAACAGGAGAGAAATGGGAC) primers. NAIP5(Δ347) (347–1402) was generated using the forward primer (AAAAGGATCCGCCACCATGACC TTGAAGTCTCTGCAGAAAG) in combination with wild-type NAIP5 reverse primer. NAIP5(ΔLRR) (1–1039) was PCR cloned into BglII/PmeI sites of MSCV2.2 vector using the wild-type NAIP5 forward primer (see above), and

reverse primer (GTTTAAACTCAGCCACTGCTGTTGAATAAACG). Wild-type NLRC4 (1–1024) was PCR amplified from pcDNA3 using forward (AAAAGGATCCGCCACCATGAACTTTATAAGGAACAACAGACG) and reverse (TTTTGCGGCCGCTTAAGCAGTCACTAGTTTAAAGGTGCC) primers, and ligated into BglII/NotI sites of MSCV2.2–IRES–GFP. NLRC4(K175R) was generated by site-directed mutagenesis using the QuickChange protocol (Stratagene) using forward (GAGTCTGGCAAAGGGCGATCGACCCTGCTGCAG) and reverse (CTGCAGCAGGGTCGATCGCCCTTGGCAGACTC) primers. NLRC4(ΔCARD) (89–1024) was PCR amplified from pcDNA3 using forward primer (AAAA GGATCCGCCACCATGTCTTATCAGGTCACAGAAGAAGACC) paired with the wild-type reverse primer. NLRC4(ΔLRR) (1–656) was PCR amplified from pcDNA3 using the reverse primer (TTTTGCGGCCGCTTACTTCCAGTTG AAGAACAAAGACACAGC) in combination with the wild-type forward primer. Wild-type NAIP2 was PCR cloned from pBluescript(SK-) into the BglII/NotI sites of MSCV2.2–IRES–GFP using forward (AAAAGGATCCGCCACCATGGCAGC CCAGGGAGAAGCCGTTGAGG) and reverse (TTTTGCGGCCGCTCACTTC TGAATGACAGGAGAGAATGGCACTACCC) primers. Wild-type NAIP6 was PCR cloned from pBluescript(SK-) into the BglII/NotI sites of Mscv2.2–IRES–GFP using the wild-type NAIP5 forward primer, and the wild-type NAIP6 reverse primer (NAIP1 and NAIP6 share identical C-terminal nucleotides). MSCV2.2–caspase 1 and MSCV2.2–caspase 1(C284A) have been described previously¹⁵.

We modified the MSCV2.2 vector to contain a 6×-Myc-tag in the MCS to facilitate generation of NH₂-terminal fusion proteins (6×-Myc–MSCV2.2–IRES–GFP). The 6×-Myc-tag from pCS-6×-Myc–SEC24 was PCR amplified using forward (AAAAAGATCTATCGATTAAAGCTATG) and reverse (TTTTGC GGCCGCTGG–CCGGCCTGAATCA) primers for insertion into the BglII site of MSCV2.2. 6×-Myc–FlaA was generated by PCR amplifying full-length FlaA (*L. pneumophila*) from MSCV2.2–FlaA⁷ using forward (AAAAGCGGCCGCGCAG CTCAAGTAATCAACACTAATGTGGC) and reverse (TTTTGTGCGACTATC–GACCTAACAAAGATAATAC) primers and inserted into NotI/SalI sites. 6×-Myc–PrgJ was generated by amplifying PrgJ from MSCV2.2–PrgJ using forward (AAAAGCGGCCGCGATCGATTGCAACTATTGTCCC) and reverse (TTTTGTGCGACTATGAGCGTAATAGCGTTTC) primers and insertion into NotI/SalI sites. All constructs were fully sequenced to confirm their identity. Sequencing primers used for NAIP5: MSCV2.2–F, AAGCCCTTTGTA CACCCTAAGCC, MSCV2.2–R, CCTCACATTGCCAAAGAG; NAIP5seq#1, CAGCAAAAGCACTGAACGCC; NAIP5seq#2, ATGAACAAATCCCTCGTAGC; NAIP5seq#3, TCACCTCTACCAAGTCCAC; NAIP5seq#4, CTCAGACAC–CTTCACTAATGC; NAIP5seq#5, TCCCTTAGTTCCATCACACC; NAIP5seq#6, GACCCCTCTCTTTGTAGCAG; NAIP5seq#7, GAGTTTCTTGCTGCCGTGAG; NAIP5seq#8, TTAGAGGGTTGT–GGCTGGTGTC; NAIP5seq#9, CTTCACA GAGTATTGAGTCCG; NAIP5seq#10, TTGAGTTTCTGAGCATGTC; NAIP5seq#11, GGCAACTTGCCAACTCAG. Sequencing primers for NAIP2: NAIP2seqF1, TGGTGATGAGAAAGAGTCAC; NAIP2seqF2, CTTCACAGAGT ATTGAGTTCG; NAIP2seqR1, AGCAAATGGTCAGTGCCGAG; NAIP2seqR2, ACATACTGCTGCCACGAAAG; NAIP2seqR3, AATCCAGTGTCTCCCTCG; MSCV2.2–F and –R primers (see above). Sequencing primers for NAIP6: NAIP6seqF1, CAGAAAGCCTGTTACTGTTAG; NAIP6seqR1, GATGGAAC TAAAGGAGAGGTAG; NAIP6seqR2, TCTTGGTCTTCTGCCTATC; MSCV2.2–F, MSCV2.2–R, NAIP5seq#7 (see above).

Generation of Naip2 shRNA constructs. We used the lentiviral pLKO.1–TRC cloning vector (Addgene) to generate three vectors expressing three separate shRNAs targeting *Naip2* (*Naip2* shRNAs 1–3). *Naip2* shRNA1 oligos, CCGGG CCATTGCCTTTCAACCTATACTCGAGTATAGGTTGAAAGGCAATGGCT TTTTG and AATTCAAAAAGCCATTGCCTTTCAACCTATACTCGAGTAT AGGTTGAAAGGCAATGGC. *Naip2* shRNA2 oligos, CCGGCCATCCAGAAA CCTTGTTGTTCTCGAGAACAAAGGTTTCTGGATGGTTTTTG and AA TTCAAAAACCATCCAGAAACCTTGTTGTTCTCGAGAACAAAGGTTT CTGGATGG. *Naip2* shRNA3 oligos, CCGGCTTTCACTCTTGAAGACAA CTCGAGTTGTCTCTTCAAGACTGAAAGTTTTG and AATTCAAAAAC TTCAGTCTTGAAGAGACAACTCGAGTTGTCTCTTCAAGACTGAAAG. We included pLKO.1–TRC control vector (Addgene ID#10879) or pLKO.1 scramble (Addgene ID#1864) as negative controls.

Knockdown of NAIP2 in primary macrophages. Bone marrow was collected from C57BL/6 mice on day 0, and plated into one 15-cm plate in macrophage differentiation media (see above). Lentivirus encoding *Naip2* shRNAs were generated according to the Addgene protocol. On day 2, primary bone marrow macrophages were collected, red blood cells were lysed, and cells were plated at 1 × 10⁶ cells per

well in 6-well plates. Macrophages were spininfected with lentiviral particles at 32 °C, 90 min, 1,258g, and placed in a 32 °C incubator for 48 h. On day 4, cells were collected in cold PBS, re-plated on 10-cm plates containing fresh media containing puromycin (5 µg ml⁻¹) for selection, and placed at 37 °C. Puromycin containing media was replaced on day 6. On day 8, macrophages were collected, counted and seeded at 1 × 10⁵ cells per well in a 96-well plate, and infected the following day. In some experiments, NAIP2 was knocked down in v-myc/v-raf immortalized bone-marrow-derived macrophages that were previously generated¹⁵ by use of the J2 virus.

Listeria infections. Flagellin-deficient strains of *Listeria monocytogenes* 10403S were generated that express the secreted fusion protein ActAN100–PrgJ (*Listeria* Δ flaA-pPrgJ) or ActAN100–FlaA (*Listeria* Δ flaA-pFlaA), under the control of an IPTG-inducible promoter, as previously described⁸. Macrophages were spininfected in antibiotic free media at 400g for 10 min at an MOI = 5, with or without IPTG

(1 mM), and placed at 37 °C for 30 min. The media was then replaced with complete media containing gentamycin (10 µg ml⁻¹) and IPTG, and supernatants were assayed for LDH assay 5.5 h later.

Salmonella infections. *Salmonella enterica* serovar Typhimurium strain LT2 or isogenic flagellin mutant (FliC⁻FljB⁻) was grown in 10 ml LB standing cultures at 37 °C overnight. The next morning, the cultures were diluted 1:100 in LB and grown for 4 h (standing culture, 37 °C). Bacteria were added to macrophages at an MOI of 10–30, followed by centrifugation at 400g for 10 min. Gentamycin (25 µg ml⁻¹) was added after 1 h to kill remaining extracellular bacteria. Caspase 1 processing or LDH release was monitored as previously described⁷.

26. Krantz, B. A. *et al.* A phenylalanine clamp catalyzes protein translocation through the anthrax toxin pore. *Science* **309**, 777–781 (2005).

The NLRC4 inflammasome receptors for bacterial flagellin and type III secretion apparatus

Yue Zhao^{1,2*}, Jieliang Yang^{1,2*}, Jianjin Shi², Yi-Nan Gong², Qiuhe Lu², Hao Xu², Liping Liu² & Feng Shao²

Inflammasomes are large cytoplasmic complexes that sense microbial infections/danger molecules and induce caspase-1 activation-dependent cytokine production and macrophage inflammatory death^{1,2}. The inflammasome assembled by the NOD-like receptor (NLR) protein NLRC4 responds to bacterial flagellin and a conserved type III secretion system (TTSS) rod component^{3–5}. How the NLRC4 inflammasome detects the two bacterial products and the molecular mechanism of NLRC4 inflammasome activation are not understood. Here we show that NAIP5, a BIR-domain NLR protein required for *Legionella pneumophila* replication in mouse macrophages⁶, is a universal component of the flagellin–NLRC4 pathway. NAIP5 directly and specifically interacted with flagellin, which determined the inflammasome-stimulation activities of different bacterial flagellins. NAIP5 engagement by flagellin promoted a physical NAIP5–NLRC4 association, rendering full reconstitution of a flagellin-responsive NLRC4 inflammasome in non-macrophage cells. The related NAIP2 functioned analogously to NAIP5, serving as a specific inflammasome receptor for TTSS rod proteins such as *Salmonella* PrgJ and *Burkholderia* BsaK. Genetic analysis of *Chromobacterium violaceum* infection revealed that the TTSS needle protein CprI can stimulate NLRC4 inflammasome activation in human macrophages. Similarly, CprI is specifically recognized by human NAIP, the sole NAIP family member in human. The finding that NAIP proteins are inflammasome receptors for bacterial flagellin and TTSS apparatus components further predicts that the remaining NAIP family members may recognize other unidentified microbial products to activate NLRC4 inflammasome-mediated innate immunity.

The NLR protein NLRC4 (also known as IPAF) in macrophages activates caspase-1 and downstream inflammatory response upon sensing cytosolic presence of flagellin during bacterial infection^{3,4,7,8}. To study the mechanism of the NLRC4 inflammasome, a defined biochemical assay was developed by fusing recombinant flagellin carboxy-terminal to the amino-terminal domain of anthrax lethal factor. This domain (designated as LFn here), through binding to another anthrax protein called protective antigen (PA), can efficiently translocate heterologous fusion proteins into mammalian cytosol through endocytosis-mediated entry⁹. Using this system, purified flagellin from *L. pneumophila* (LFn-FlaA^{LP}) was found to trigger robust caspase-1 cleavage (Fig. 1a), IL-1 β release and pyroptotic death (Supplementary Fig. 1a, b) in primary bone-marrow-derived macrophages (BMMs). These activations were completely diminished in *Nlrp4*^{−/−} and caspase-1^{−/−} macrophages (Fig. 1a and Supplementary Fig. 1a, b); *Asc*^{−/−} (*Asc* also known as *Pycard*) BMMs also showed little caspase-1 maturation and IL-1 β release but with a partially affected pyroptosis due to ASC-independent NLRC4 inflammasome activation¹⁰. Full activation of NLRC4 inflammasome requires L470, L472 and L473 in *Legionella* flagellin⁶. Accordingly, alanine substitutions of the three leucine residues (3A) generated a largely inactive LFn-FlaA^{LP} protein (Fig. 1a and Supplementary Fig. 1a, b). LFn-FlaA^{LP} induced similar NLRC4-dependent caspase-1 activation and pyroptosis in

Tlr4^{−/−} macrophages (Supplementary Figs 2a, b and 3a), excluding a possible contribution from residual endotoxin contaminants present in the recombinant protein. Other bacteria such as *Salmonella typhimurium* also trigger flagellin-dependent NLRC4 inflammasome activation^{3,4,8}. Delivery of *S. typhimurium* (LFn-FliCSt) or *Yersinia enterocolitica* (LFn-FliC2^{Ye}) flagellin into BMMs induced robust caspase-1 activation and extensive pyroptosis in an NLRC4-dependent manner (Supplementary Fig. 4a, b). Thus, LFn-mediated delivery of recombinant flagellin recapitulates all known genetic properties of flagellin activation of the NLRC4 inflammasome.

For *L. pneumophila* infection, flagellin-induced caspase-1 activation requires NAIP5 (also known as BIRC1E), a BIR-domain-containing NLR protein⁶. A natural variant of NAIP5 renders macrophages from the A/J mouse permissive to *L. pneumophila* intracellular replication^{11–16}. The role of NAIP5 for other bacterial flagellins is not clear^{6,17}. RNA interference (RNAi) knockdown of *Naip5* (Supplementary Fig. 3b) severely blocked LFn-FlaA^{LP}-triggered caspase-1 activation and pyroptosis (Supplementary Fig. 2a, b). Notably, activation of the NLRC4 inflammasome by LFn-FliCSt and LFn-FliC2^{Ye}, but not that by *Salmonella* TTSS rod protein (LFn-PrgJ), was also drastically reduced by short hairpin RNA (shRNA)-mediated stable knockdown of *Naip5* (Fig. 1b, c and Supplementary Fig. 3c). Consistently, flagellin-triggered caspase-1 activation during *Salmonella* and *Legionella* infection was significantly attenuated in *Naip5* knockdown macrophages (Fig. 1d).

The finding that NAIP5 is a possible integral component of the flagellin–NLRC4 pathway inspired us to investigate whether NAIP5 directly recognizes flagellin. *Legionella* flagellin was found to show an evident yeast two-hybrid interaction with NAIP5, but not NLRC4, whereas the 3A mutant showed no interaction (Fig. 2a). *Naip5* is located within a genomic locus containing seven highly homologous *Naip* genes (*Naip1–7*) and four of them (*Naip1*, *Naip2*, *Naip5* and *Naip6*) have transcripts in C57BL/6 mice¹¹. *Legionella* flagellin also showed a two-hybrid interaction with NAIP6, but not with NAIP1 and NAIP2 (Fig. 2a). Supporting the two-hybrid results, *Legionella* flagellin expressed in 293T cells readily co-precipitated NAIP5 and NAIP6, but not NAIP1, NAIP2 and NLRC4, whereas the 3A mutant failed to do so (Fig. 2b). The TLR5-binding-deficient mutant (I391A), which is fully functional in inflammasome activation⁶, behaved similarly to wild-type flagellin in the co-immunoprecipitation assay. Flagellin also co-precipitated NAIP5 encoded by the A/J allele (NAIP5^{A/J}) (Fig. 2b), which explains the normal or nearly normal caspase-1 activation in *L. pneumophila*-infected A/J macrophages^{15,18,19}.

A panel of nine additional flagellins from different bacteria was further profiled (Fig. 2c). In the two-hybrid assay, flagellins from *S. typhimurium*, *Y. enterocolitica*, *Photobacterium luminescens* and *Pseudomonas aeruginosa* showed a positive result whereas those from enteropathogenic *Escherichia coli* (EPEC), enterohaemorrhagic *E. coli* (EHEC), *Shigella flexneri*, *Chromobacterium violaceum* and *Burkholderia thailandensis* did not interact with NAIP5 (Fig. 2c and Supplementary Fig. 5a). NAIP5 interaction with *S. typhimurium* flagellin

¹Graduate Program in Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100730, China. ²National Institute of Biological Sciences, Beijing, 102206, China.

*These authors contributed equally to this work.

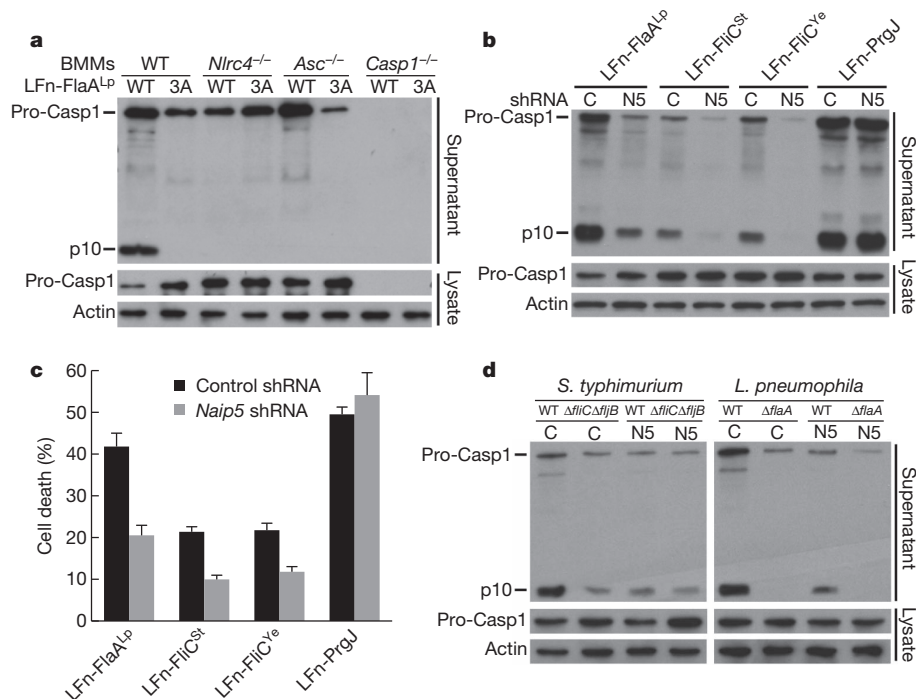


Figure 1 | A defined biochemical assay reveals a universal role of NAIP5 in flagellin-triggered NLRC4 inflammasome activation in mouse macrophages. **a**, Effects of anthrax lethal factor N-terminal-domain-mediated intracellular delivery of *Legionella* flagellin (LFn-FlaA^{LP}) on caspase-1 activation in lipopolysaccharide (LPS)-primed BMMs derived from wild-type (WT, C57BL/6) or indicated knockout mice. 3A denotes a triple mutant flagellin (L470A/L472A/L473A). Shown are anti-caspase-1 and anti-actin immunoblots of culture supernatants (top) and total cell lysates (bottom). p10 denotes the processed mature form of caspase-1. **b**, **c**, Effects of *Naip5* knockdown on flagellin-induced caspase-1 activation (**b**) and cell death (**c**). **d**

Naip5-targeting (N5) (Supplementary Table 1) or a control (C) shRNA was stably expressed in immortalized BMMs. LFn-FlaA^{LP}, FliCSt and FliC^{Ye} are recombinant LFn-tagged flagellins from *L. pneumophila*, *S. typhimurium* and *Y. enterocolitica*, respectively. LFn-PrgJ is LFn-tagged *S. typhimurium* TTSS rod protein. **c**, LDH releases are shown as mean values \pm standard deviation (s.d.) from three independent determinations. **d**, Effects of *Naip5* knockdown on flagellin-induced caspase-1 activation during *Salmonella* and *Legionella* infection. Δ fliC Δ fliB and Δ fliA denote flagellin-deficient strains of *S. typhimurium* and *L. pneumophila*, respectively.

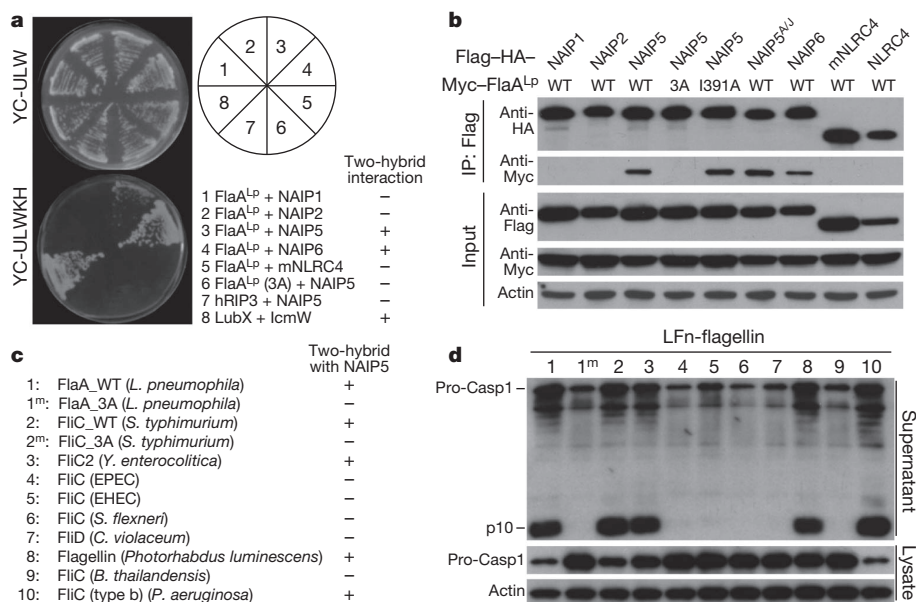


Figure 2 | Flagellin interacts specifically with NAIP5 and the interaction correlates with the activity of flagellins from different bacteria. **a**, Yeast two-hybrid interaction assays of *Legionella* flagellin (FlaA^{LP}) and different NAIP proteins (or mouse (m)NLRC4). The chart in the lower right corner summarizes the interaction results. The known interaction between *Legionella* effector LubX and its secretion chaperon IcmW was included as a positive control. **b**, Co-immunoprecipitation assays of *Legionella* flagellin (FlaA^{LP}) and

different NAIP proteins (or NLRC4). Shown are immunoblots of anti-Flag immunoprecipitates (IP: Flag) and total cell lysates (Input). I391A is a TLR5 binding-deficient flagellin mutant. **c**, Summary of yeast two-hybrid interaction of NAIP5 with different bacterial flagellins. The raw data are shown in Supplementary Fig. 5a. **d**, Caspase-1 activation assays of LFn-mediated delivery of different bacterial flagellins into primary BMMs. Number denotations of different flagellins follow those in **c**.

required three leucine residues equivalent to those in *Legionella* flagellin. When delivered into BMMs, flagellins from *S. typhimurium*, *Y. enterocolitica*, *P. luminescens* and *P. aeruginosa*, but not those from the other five bacteria species, stimulated caspase-1 activation, macrophage death and IL-1 β release (Fig. 2d and Supplementary Fig. 5b, c). The positive NAIP5-binding and inflammasome-stimulating activities of *S. typhimurium* and *P. aeruginosa* flagellins agree with their genetic requirements for infection-induced caspase-1 activation^{3,4,8,20–22}. Among those inactive ones, *S. flexneri* flagellin is not expressed and dispensable for host innate immune detection of *S. flexneri* infection²³. Genetic ablations of flagellins from EPEC and *B. thailandensis* also did not affect infection-induced caspase-1 activation (Supplementary Fig. 6). Thus, the ability of the ten different flagellins to interact with NAIP5 correlates well with their differential inflammasome-stimulating activity, which further supports the idea that flagellin is generally recognized by NAIP5 in triggering NLRC4 inflammasome activation.

NAIP5 and NLRC4 were then co-expressed in 293T cells and their possible interactions were investigated. Co-immunoprecipitation of NAIP5 and NLRC4 was barely detectable in the absence of flagellin. However, co-expression of *Legionella* flagellin, but not the 3A mutant, significantly increased the amount of NAIP5 precipitated by NLRC4 (Fig. 3a, b). Flagellin was also detected in the NLRC4 immunoprecipitates due to the bridging effect of NAIP5. Deletion of the nucleotide-binding P-loop in NLRC4 nucleotide-binding and oligomerization domain (NOD) abolished flagellin-stimulated NLRC4–NAIP5 interaction (Fig. 3a), which agrees with the reported interaction between NOD domains from NLRC4 and NAIP5 (ref. 24). Flagellin also promoted the association of NLRC4 with NAIP5^{A/J}, but neither NAIP1 nor NAIP2 was precipitated by NLRC4 despite the presence of flagellin (Fig. 3b). To test whether the flagellin-stimulated NAIP5–NLRC4 complex can activate downstream signalling, NAIP5 and NLRC4, together with pro-caspase-1 and pro-IL-1 β , were co-expressed in 293T cells. Delivery of LFn-FlaA^{LP}, but not the 3A mutant, into the transfected cells resulted in an evident production of mature IL-1 β (Fig. 3c and Supplementary Fig. 7a). Omission of NAIP5, NLRC4 or caspase-1 in this reconstitution abolished the response to flagellin

stimulation. NAIP5^{A/J} also supported the reconstitution whereas NAIP1 and NAIP2 failed to do so (Fig. 3d and Supplementary Fig. 7b), consistent with their differential association with NLRC4 upon flagellin stimulation (Fig. 3b). Moreover, the reconstituted NLRC4 inflammasome exhibited robust responses to flagellins from *Salmonella* and *Yersinia* (Supplementary Fig. 8a). Each of the three domains in both NLR proteins (CARD, NOD and LRR in NLRC4; BIR, NOD and LRR in NAIP5) was essential for assembling a flagellin-responsive inflammasome complex (Supplementary Fig. 8b). These results indicate that flagellin recognition by NAIP5 stimulates the physical association between NAIP5 and NLRC4, thereby signalling downstream caspase-1 activation.

NAIP6 interacted with flagellin in a manner similar to NAIP5 (Fig. 2b and Supplementary Fig. 9) and supported the reconstitution in 293T cells (Fig. 3b, d). In fact, NAIP6, among all the NAIP proteins, shares the highest sequence identity with NAIP5, of 94.7% (Supplementary Fig. 10). NAIP6 probably has a similar function to NAIP5 in macrophage detection of flagellin, but its role might be relatively minor given its much lower expression in primary macrophages compared with that of NAIP5 (ref. 12).

The NLRC4 inflammasome also responds to a conserved TTSS rod protein such as PrgJ in *S. typhimurium*, BsaK in *B. thailandensis* and EscI in EPEC⁵. Delivery of recombinant BsaK (LFn-BsaK) into BMMs recapitulated such effects and induced NLRC4-dependent caspase-1 activation and pyroptosis (Supplementary Fig. 11). Given that NAIP5 recognizes flagellin and that PrgJ activation of the NLRC4 inflammasome is independent of NAIP5 (ref. 17), we proposed that other NAIP proteins could recognize the TTSS rod protein. BsaK was found to interact with NAIP2, but not NAIP1, NAIP5, NAIP6 and NLRC4, in the two-hybrid assay (Fig. 4a). Co-immunoprecipitation assay confirmed this NAIP2-specific interaction (Fig. 4b). This observation agrees with the idea that NAIP2 is the most distantly related to the other NAIP proteins (Supplementary Fig. 10). Reconstitution in non-macrophage cells further showed that only NAIP2, but not any other NAIP, effected robust IL-1 β maturation upon LFn-BsaK stimulation (Fig. 4c). These findings indicate that NAIP2 is the specific receptor for the TTSS rod protein.

To test the requirement of NAIP2 for detecting the TTSS rod protein in macrophages, *Naip2* stable knockdown BMMs were generated. Among the four different shRNAs (*Naip2-1*, 2, 3 and 4), *Naip2-1* and *Naip2-2* considerably reduced *Naip2* messenger RNA level whereas *Naip2-3* and *Naip2-4* showed intermediate and negligible efficiency, respectively (Supplementary Fig. 12a). *Naip2-1* and *Naip2-2* knockdown macrophages exhibited significant resistance in caspase-1 activation and pyroptosis to LFn-PrgJ or LFn-BsaK stimulation (Supplementary Fig. 12b–d). In contrast, *Naip2-3* knockdown macrophages showed a mild resistance and *Naip2-4* knockdown macrophages had a normal sensitivity to rod protein stimulations. In *Naip2-2* knockdown macrophages, in which mRNA levels of other *Naip* genes were not affected (Supplementary Fig. 13), attenuated caspase-1 activation was only observed with the rod protein stimulations, but not with flagellin stimulations (Fig. 4d). Furthermore, deletion of genes encoding the rod proteins from flagellin-deficient EPEC and *S. typhimurium* abolished bacterial infection-induced caspase-1 activation, and this effect did not occur in *Naip2-2* knockdown macrophages (Fig. 4e). These results demonstrate the critical and specific role of NAIP2 in recognizing the TTSS rod protein for NLRC4 inflammasome activation.

In contrast to mouse macrophages, human U937 monocyte-derived macrophages were unresponsive to intracellular delivery of flagellin and BsaK/PrgJ-like rod protein (Supplementary Fig. 14). When profiling our genetic collection of various pathogenic bacteria, a *C. violaceum* strain (deficient in secretion of TTSS effectors) was identified to be capable of inducing caspase-1 activation in human U937 monocytes (Supplementary Fig. 15a). Notably, further ablation of five possible flagellin genes (ΔF) caused no reduction in this activation. Stable knockdown of *NLRC4* (Supplementary Fig. 16) significantly attenuated *C.*

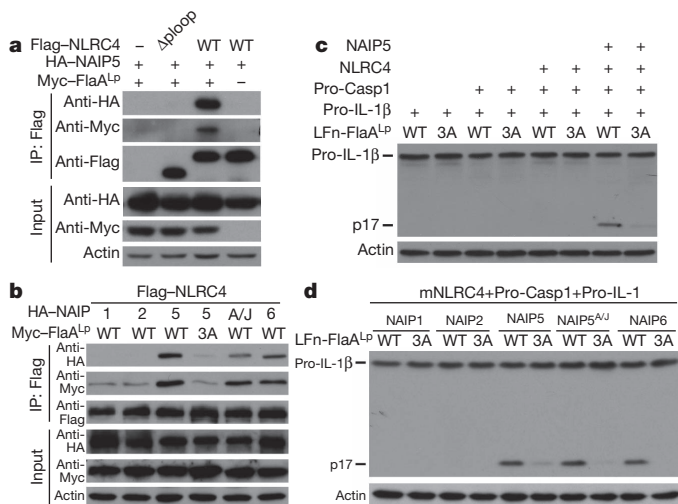


Figure 3 | Flagellin stimulates the NAIP5–NLRC4 association and reconstitution of flagellin activation of the NLRC4 inflammasome in non-macrophage cells. **a, b**, Co-immunoprecipitation assays of NAIP5 and NLRC4 interaction in the presence or absence of flagellin. Δ ploop in **a** denotes an NLRC4 mutant with deletion of the nucleotide-binding P-loop. **c**, Reconstitution of flagellin activation of the NLRC4 inflammasome in non-macrophage cells. Lysates from 293T cells transfected with indicated plasmid combinations and stimulated with LFn-FlaA^{LP} were analysed for mature IL-1 β (p17) by immunoblotting. Expression of transfected inflammasome components for **c** and **d** is in Supplementary Fig. 7. **d**, Assay of different NAIP proteins in supporting reconstitution of flagellin activation of the NLRC4 inflammasome in 293T cells.

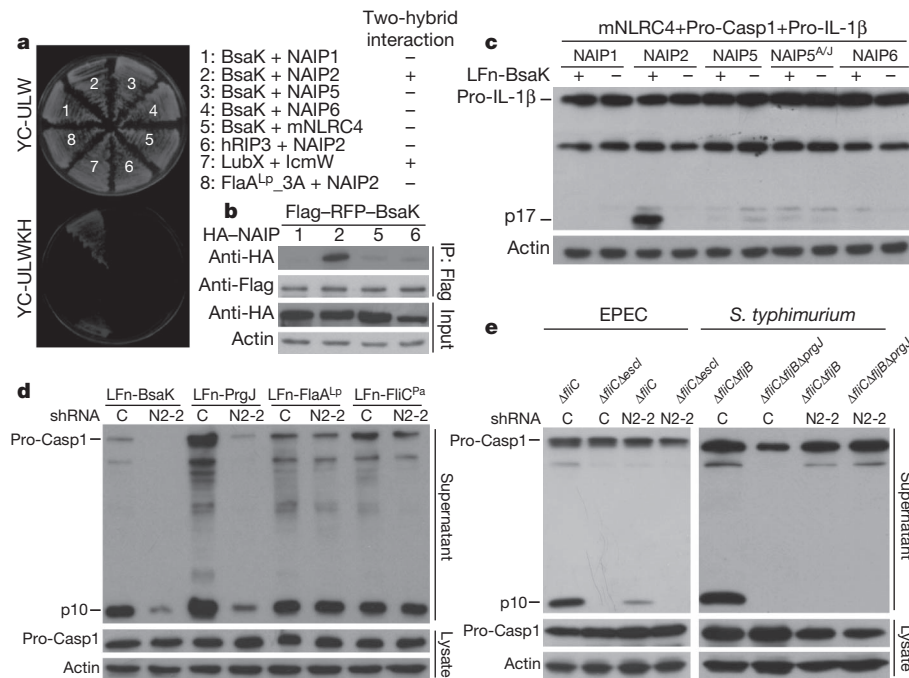


Figure 4 | NAIP2 interacts with the TTSS rod protein and is required for the rod protein to trigger mouse NLRC4 inflammasome activation. **a**, Yeast two-hybrid (**a**) and co-immunoprecipitation (**b**) assays of interactions between *B. thailandensis* rod protein BsaK and different NAIP proteins. **c**, Reconstitution of BsaK activation of the NLRC4 inflammasome in non-macrophage cells. Lysates from HeLa cells transfected with indicated plasmid combinations and stimulated with LFn-BsaK were analysed for mature IL-1 β (p17) by immunoblotting. Expression of transfected inflammasome components is in Supplementary Fig. 7. **d**, Effects of *Naip2* knockdown on

caspase-1 activation induced by TTSS rod proteins and flagellins. Control (C) or *Naip2*-2 (N2-2) stable knockdown macrophages (Supplementary Fig. 12) were stimulated with purified LFn-tagged BsaK, PrgJ, FlaA^{LP} or FliC^{Pa} proteins as indicated. **e**, Effects of *Naip2* knockdown on rod-protein-induced caspase-1 activation during EPEC and *Salmonella* infection. EPEC E2348/69 Δ fliC Δ escI and *S. typhimurium* Δ fliC Δ escI denote the rod-protein-deficient EPEC and *S. typhimurium* strains, respectively, which were constructed on the flagellin-deletion background.

violaceum Δ F-triggered caspase-1 activation and pyroptosis (Fig. 5a and Supplementary Fig. 17a). The PrgJ homologue in the *C. violaceum* Cpi-1 TTSS system, CprI, is encoded in a separate Cpi-1a locus that harbours several additional TTSS apparatus genes²⁵ (Fig. 5b). Although

cprI was not required for infection-induced caspase-1 activation and pyroptosis, deletion of the entire Cpi-1a locus largely diminished *C. violaceum*-induced NLRC4 inflammasome activation (Fig. 5a, Supplementary Fig. 15b and Supplementary Fig. 17). Further genetic analysis of the entire Cpi-1a locus identified *cprI*, which was essential for inducing caspase-1 activation and pyroptosis (Fig. 5c and Supplementary Fig. 17b). A CprI-expressing plasmid could rescue the deficiencies of inflammasome activation for both *cprI* and Cpi-1a deletion strains (Fig. 5d). Thus, *C. violaceum* requires *cprI* to stimulate NLRC4 inflammasome activation in human macrophages.

cprI encodes the conserved TTSS needle subunit that is a sequence paralogue of the rod protein²⁶, raising a hypothesis that the needle protein is the bacterial ligand recognized by the human NLRC4 inflammasome. Consistent with the above genetic analyses, LFn-mediated delivery of CprI, but not other Cpi-1a-encoded proteins, triggered robust caspase-1 activation and pyroptosis in U937 cells

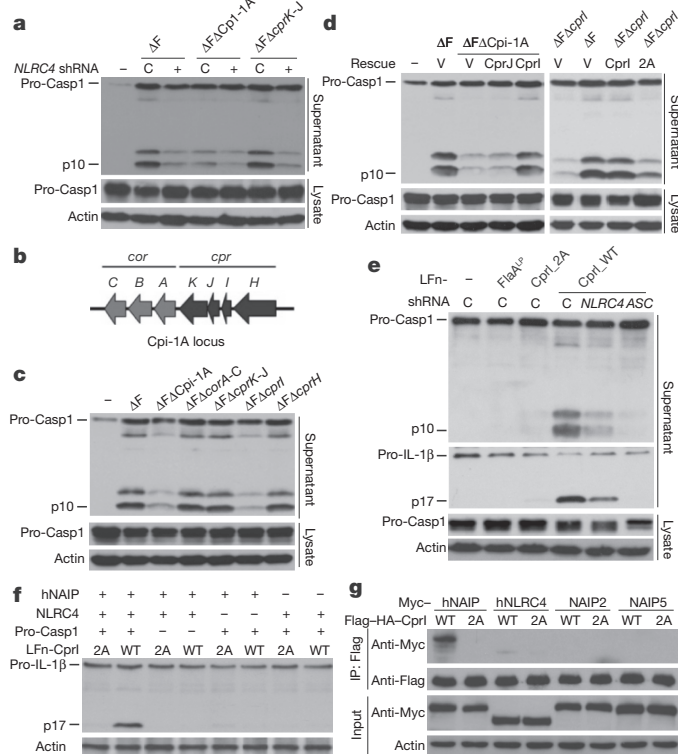


Figure 5 | *C. violaceum* infection studies reveal that the human NLRC4 inflammasome responds to the TTSS needle subunit through specific recognition by human NAIP. **a–c**, Caspase-1 activation assays of *C. violaceum* infections of human U937 monocyte-derive macrophages. Δ F has deletions of five possible flagellin genes in *C. violaceum*. Control (C) or NLRC4 (+) stable knockdown cells were used in **a**. Δ F Δ Cpi-1A means a deletion of the entire TTSS Cpi-1A locus illustrated in the schematic drawing shown in **b**. Detailed information for all the mutant strains are listed in Supplementary Table 3. **d**, Complementation of Cpi-1a locus or *cprI* deletion *C. violaceum* strains by a CprI-expressing plasmid. PMA-differentiated U937 cells were infected with indicated *C. violaceum* mutant or rescue strain. 2A is a double mutant of CprI (V69A/I79A). **e**, Caspase-1 activation assays of delivery of CprI into human U937 macrophages and effects of NLRC4 and ASC knockdown. Control (C) or NLRC4 or ASC stable knockdown cells were stimulated with LFn-CprI or other indicated LFn fusion proteins. **f**, Reconstitution of CprI activation of the human NLRC4 inflammasome in 293T cells. **g**, Co-immunoprecipitation assay of CprI and different NAIP proteins (or NLRC4).

(Supplementary Fig. 18a, b), which were largely decreased in *NLRC4* and ASC knockdown cells (Fig. 5e and Supplementary Fig. 18c). Mutation of two hydrophobic residues (V69A/I79A, 2A) in a helical hairpin region in CprI diminished its activity of stimulating inflammasome activation (Fig. 5d, e and Supplementary Fig. 18). CprI activation of the NLRC4 inflammasome could also be robustly reconstituted in 293T cells and the 2A mutant remained inactive in this assay (Fig. 5f). Most importantly, this reconstitution required human NAIP, the sole NAIP family member in human. Human NAIP-based reconstitution specifically responded to LFn-CprI, but not to LFn-FlaA^{LP} and LFn-BsaK; LFn-CprI did not activate NAIP5- and NAIP2-based reconstitution (Supplementary Fig. 19). Furthermore, CprI readily co-precipitated human NAIP, but not any of NAIP2, NAIP5 and NLRC4 from 293T cells, and the nonfunctional 2A mutant failed to interact with human NAIP (Fig. 5g). Homologous needle subunits from EHEC, *B. thailandensis*, *P. aeruginosa*, *S. flexneri* and *S. typhimurium*, but not those from EPEC and *V. parahaemolyticus*, also stimulated NLRC4 inflammasome activation in U937 cells (Supplementary Fig. 20). Thus, human NAIP functions analogously to mouse NAIP5/2, but specifically recognizes the TTSS needle subunit to trigger human NLRC4 inflammasome activation.

In summary, murine NLR proteins NAIP5 and NAIP2 directly recognize bacterial flagellin and TTSS rod protein, respectively, whereas human NAIP serves as a specific receptor for the TTSS needle protein. Engagement of NAIP receptors by corresponding bacterial ligands promotes their physical association with NLRC4, resulting in activation of the NLRC4 inflammasome and macrophage innate immunity. The inflammasome-stimulating activities of flagellin, TTSS rod and needle proteins lie in their C-terminal leucine-rich helical hairpin regions that share structural commonalities^{5,27}. Thus, other homologous NAIP proteins might recognize additional bacterial products of similar biochemical features for counteracting diverse bacterial infections. Our results also indicate that NLRC4 acts as an adaptor through which inflammasome activation signals generated from different NAIP receptors are transduced to caspase-1. Involvement of an additional cytosolic pattern recognition receptor (PRR) protein for sensing one microbial product has previously been noted with NLRP3 and NALP1-mediated inflammasome activation^{28,29}. Future studies will probably identify more PRR proteins that act sequentially within a single inflammasome complex in response to microbial products or danger signals.

METHODS SUMMARY

LFn-mediated intracellular delivery and RNAi. For delivery into macrophages, purified recombinant proteins were washed with 60% isopropanol to remove the majority of endotoxin contaminants. LFn-flagellin, LFn-BsaK/PrgI/CprI, LFn-CprI or other indicated control proteins together with PA proteins were added into culture medium (serum-free) at a final concentration of 1 µg ml⁻¹ for each protein. Cells were further incubated for 1 h (primary BMMs) or 3 h (immortalized BMMs) before being subjected to the indicated inflammasome activation assays. Transient small interfering RNA (siRNA) knockdown in macrophages was performed using the INTERFERin reagent (Polyplus Transfection) by following the manufacturer's instruction. To achieve stable knockdown in macrophages, a modified pLKO.1-GFP plasmid harbouring a specific shRNA (Supplementary Table 1) was transduced into BMMs or U937 cells by lentiviral infection and GFP-positive knockdown cells were sorted out by flow cytometry for further functional analysis.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 30 May; accepted 30 August 2011.

Published online 14 September 2011.

1. Lamkanfi, M. & Dixit, V. M. Inflammasomes: guardians of cytosolic sanctity. *Immunol. Rev.* **227**, 95–105 (2009).
2. Schroder, K. & Tschopp, J. The inflammasomes. *Cell* **140**, 821–832 (2010).
3. Franchi, L. *et al.* Cytosolic flagellin requires Ipaf for activation of caspase-1 and interleukin 1β in salmonella-infected macrophages. *Nature Immunol.* **7**, 576–582 (2006).
4. Miao, E. A. *et al.* Cytoplasmic flagellin activates caspase-1 and secretion of interleukin 1β via Ipaf. *Nature Immunol.* **7**, 569–575 (2006).

5. Miao, E. A. *et al.* Innate immune detection of the type III secretion apparatus through the NLRC4 inflammasome. *Proc. Natl Acad. Sci. USA* **107**, 3076–3080 (2010).
6. Lightfield, K. L. *et al.* Critical function for Naip5 in inflammasome activation by a conserved carboxy-terminal domain of flagellin. *Nature Immunol.* **9**, 1171–1178 (2008).
7. Amer, A. *et al.* Regulation of Legionella phagosome maturation and infection through flagellin and host Ipaf. *J. Biol. Chem.* **281**, 35217–35223 (2006).
8. Broz, P. *et al.* Redundant roles for inflammasome receptors NLRP3 and NLRC4 in host defense against *Salmonella*. *J. Exp. Med.* **207**, 1745–1755 (2010).
9. Milne, J. C., Blanke, S. R., Hanna, P. C. & Collier, R. J. Protective antigen-binding domain of anthrax lethal factor mediates translocation of a heterologous protein fused to its amino- or carboxy-terminus. *Mol. Microbiol.* **15**, 661–666 (1995).
10. Broz, P., von Moltke, J., Jones, J. W., Vance, R. E. & Monack, D. M. Differential requirement for Caspase-1 autoproteolysis in pathogen-induced cell death and cytokine processing. *Cell Host Microbe* **8**, 471–483 (2010).
11. Diez, E. *et al.* *Birc1e* is the gene within the *Lgn1* locus associated with resistance to *Legionella pneumophila*. *Nature Genet.* **33**, 55–60 (2003).
12. Wright, E. K. *et al.* Naip5 affects host susceptibility to the intracellular pathogen *Legionella pneumophila*. *Curr. Biol.* **13**, 27–36 (2003).
13. Molofsky, A. B. *et al.* Cytosolic recognition of flagellin by mouse macrophages restricts *Legionella pneumophila* infection. *J. Exp. Med.* **203**, 1093–1104 (2006).
14. Ren, T., Zamboni, D. S., Roy, C. R., Dietrich, W. F. & Vance, R. E. Flagellin-deficient *Legionella* mutants evade caspase-1- and Naip5-mediated macrophage immunity. *PLoS Pathog.* **2**, e18 (2006).
15. Zamboni, D. S. *et al.* The *Birc1e* cytosolic pattern-recognition receptor contributes to the detection and control of *Legionella pneumophila* infection. *Nature Immunol.* **7**, 318–325 (2006).
16. Fortier, A., de Chastellier, C., Balor, S. & Gros, P. *Birc1e/Naip5* rapidly antagonizes modulation of phagosome maturation by *Legionella pneumophila*. *Cell. Microbiol.* **9**, 910–923 (2007).
17. Lightfield, K. L. *et al.* Differential requirements for NAIP5 in activation of the NLRC4 inflammasome. *Infect. Immun.* **79**, 1606–1614 (2011).
18. Lamkanfi, M. *et al.* The Nod-like receptor family member Naip5/Birc1e restricts *Legionella pneumophila* growth independently of caspase-1 activation. *J. Immunol.* **178**, 8022–8027 (2007).
19. Akhter, A. *et al.* Caspase-7 activation by the Nlr4/Ipaf inflammasome restricts *Legionella pneumophila* infection. *PLoS Pathog.* **5**, e1000361 (2009).
20. Franchi, L. *et al.* Critical role for Ipaf in *Pseudomonas aeruginosa*-induced caspase-1 activation. *Eur. J. Immunol.* **37**, 3030–3039 (2007).
21. Sutterwala, F. S. *et al.* Immune recognition of *Pseudomonas aeruginosa* mediated by the IPAF/NLRC4 inflammasome. *J. Exp. Med.* **204**, 3235–3245 (2007).
22. Miao, E. A., Ernst, R. K., Dors, M., Mao, D. P. & Aderem, A. *Pseudomonas aeruginosa* activates caspase 1 through Ipaf. *Proc. Natl Acad. Sci. USA* **105**, 2562–2567 (2008).
23. Suzuki, T. *et al.* Differential regulation of caspase-1 activation, pyroptosis, and autophagy via Ipaf and ASC in *Shigella*-infected macrophages. *PLoS Pathog.* **3**, e111 (2007).
24. Damiano, J. S., Oliveira, V., Welsh, K. & Reed, J. C. Heterotypic interactions among NACHT domains: implications for regulation of innate immune responses. *Biochem. J.* **381**, 213–219 (2004).
25. Miki, T. *et al.* Chromobacterium pathogenicity island 1 type III secretion system is a major virulence determinant for *Chromobacterium violaceum*-induced cell death in hepatocytes. *Mol. Microbiol.* **77**, 855–872 (2010).
26. Worrall, L. J., Lameignere, E. & Strynadka, N. C. Structural overview of the bacterial injectosome. *Curr. Opin. Microbiol.* **14**, 3–8 (2011).
27. Poyraz, O. *et al.* Protein refolding is required for assembly of the type three secretion needle. *Nature Struct. Mol. Biol.* **17**, 788–792 (2010).
28. Hsu, L. C. *et al.* A NOD2–NALP1 complex mediates caspase-1-dependent IL-1β secretion in response to *Bacillus anthracis* infection and muramyl dipeptide. *Proc. Natl Acad. Sci. USA* **105**, 7803–7808 (2008).
29. Poeck, H. *et al.* Recognition of RNA virus by RIG-I results in activation of CARD9 and inflammasome signaling for interleukin 1β production. *Nature Immunol.* **11**, 63–69 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank V. Dixit for providing *Nlr4* and *Asc* knockout mice, K. Fitzgerald, D. Radzioch and A. Ding for immortalized macrophages, R. Vance for *Naip5^{ΔJ}* cDNA, M. Donnenberg and J. Girón for EPEC strains, E. Miao for flagellin-deficient *S. typhimurium* strain, D. Milton and T. Hoang for bacterial vectors and T. Miki for *C. violaceum* strains. We are grateful to C. Yao for helping with flow cytometry, and Y. Xu and the NIBS animal facility for handling mouse lines. We thank members of the F.S. laboratory for helpful discussions and technical assistance. This work was supported by the National Basic Research Program of China (973 Programs, 2010CB835400 and 2012CB518700).

Author Contributions Y.Z. and J.Y. performed experiments, assisted by J.S., Y.-N.G., Q.L., H.X. and L.L. Y.Z., J.Y. and F.S. analysed the data and wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to F.S. (shaofeng@nibs.ac.cn).

METHODS

Plasmids, antibodies and reagents. DNAs for flagellin were amplified from the corresponding bacterial genomic DNA, and cloned into pET28a-LFn vector (Addgene) for recombinant expression in *E. coli* as described previously^{30,31}. BsaK and PrgJ DNAs were amplified from *B. thailandensis* E264 and *S. typhimurium* LT2 strains, respectively, and inserted into the same pET28a-LFn vector. DNAs for CprI, CprJ, CorB and CorC were amplified from *C. violaceum* strain (ATCC accession 12472) and also cloned into pET28a-LFn vector to prepare recombinant LFn fusion protein. PA expression plasmid was also obtained from Addgene. To construct the complementation plasmid for the *C. violaceum* mutant, CprI or CprJ DNAs with ribosome binding site (RBS) sequence were cloned into the pBBR1MCS2 vector. Expression plasmids for pro-caspase-1 and pro-IL-1 β were provided by X. Wang (University of Texas Southwestern Medical Center). cDNAs for mouse NAIP1, NAIP2, NAIP5^{C57BL/6}, NAIP6, human NAIP and NLRC4 were amplified from IMAGE EST clones (40130690, 40086453, 6850660, 100068362, 9052275 and 5179909, respectively) and mouse NLRC4 was amplified from reverse-transcribed mouse cDNA. For mammalian expression, cDNAs for all NLR proteins were cloned into modified pCS2 vectors with an N-terminal Myc, HA or Flag epitope tag. All truncations and point mutations were generated by standard molecular biology procedures. All plasmids were verified by DNA sequencing.

Antibodies for caspase-1 and Myc epitopes were obtained from Santa Cruz Biotechnology. Other antibodies used in this study include IL-1 β (3ZD; Biological Resources Branch, National Cancer Institute), HA epitope (Covance) and Flag M2 (Sigma). 293T and HeLa cells obtained from ATCC were grown in Dulbecco's modified Eagle's medium containing 10% fetal bovine serum and 2 mM L-glutamine at 37 °C in a 5% CO₂ incubator. Cell culture products were from Invitrogen and all other chemicals were Sigma-Aldrich products unless noted.

Mouse BMMs and human monocyte-derived macrophages. C57BL/6 wild-type mice were from Vital River Laboratory Animal Technology Co. and caspase-1^{-/-} mice³² were obtained from the Jackson Laboratory. *Nlrc4*^{-/-} and *Asc*^{-/-} mice³³ were provided by V. Dixit (Genentech). All knockout alleles have been crossed onto the C57BL/6 background. All animal experiments were conducted following the Ministry of Health national guidelines for housing and care of laboratory animals and performed in accordance with institutional regulations after review and approval by the Institutional Animal Care and Use Committee at National Institute of Biological Sciences. Primary BMMs were prepared by following a standard procedure as previously described³⁴. An immortalized macrophage line derived from C57BL/6 mice was provided by K. A. Fitzgerald (University of Massachusetts Medical School) and TLR4-deficient immortalized BMMs was a gift from A. Ding (Cornell University). Human U937 monocytes obtained from ATCC were cultured in RPMI-1640 containing 10% FBS and 2 mM L-glutamine and grown at 37 °C with 5% CO₂. 50 ng ml⁻¹ PMA was used to induce U937 differentiation for 48 h. Differentiated U937 cells were digested with 2 mM EDTA in PBS and subcultured in 24-well plates for further experiment.

Yeast two-hybrid and co-immunoprecipitation assays. Indicated flagellin, *bsaK* and *prgJ* genes were cloned into the bait vector pLexAde, and mouse *Naip1*, *Naip2*, *Naip5*, *Naip6* cDNAs and *Nlrc4* cDNAs were cloned into the prey vector pVP16. The bait and prey plasmids were co-transformed into the reporter *Saccharomyces cerevisiae* strain L40 by using the lithium acetate method. Two-hybrid assays were performed by following a classical procedure³⁵.

For immunoprecipitation, 293T cells were transfected with indicated plasmids. Cells were harvested and lysed in a buffer containing 50 mM Tris-HCl (pH 7.6), 150 mM NaCl and 1% Triton X-100 supplemented with a protease inhibitor mixture (Roche Molecular Biochemicals). Precleared lysates were subjected to anti-Flag M2 immunoprecipitation by following the manufacturer's instructions. The beads were washed three times with the lysis buffer and the immunoprecipitates were eluted in the SDS sample buffer followed by immunoblotting analysis. All the immunoprecipitation assays were performed more than three times and representative results are shown in the figures.

Purification of recombinant proteins. *E. coli* BL21 (DE3) strains harbouring the expression plasmids were grown in Luria-Bertani medium (tryptone, 10 g l⁻¹, yeast extract, 5 g l⁻¹, NaCl, 10.0 g l⁻¹) supplemented with appropriate antibiotics. Protein expression was induced overnight at 22 °C with 0.4 mM isopropyl- β -D-thiogalactopyranoside (IPTG) after OD_{600 nm} reached 0.8. Bacteria were harvested and lysed in a buffer containing 50 mM Tris-HCl (pH 7.6), 300 mM NaCl and 25 mM imidazole. His-tagged proteins were purified by affinity chromatography using Ni-NTA beads (Qiagen). To remove the majority of endotoxin contaminants, proteins bound onto the Ni-NTA column were subjected to an additional wash with 60% isopropanol in the wash buffer (>30 \times column volume). Proteins were then eluted with 250 mM imidazole in 50 mM Tris-HCl (pH 7.6) and 300 mM NaCl. Eluted samples were further dialysed against a buffer containing 50 mM Tris-HCl (pH 7.6) and 150 mM NaCl to remove the imidazole. Protein

concentrations were estimated by Coomassie blue staining of SDS-PAGE gels using BSA as the standards.

NLRC4 inflammasome reconstitution in HeLa and 293T cells. For reconstitution in 293T or HeLa cells, cells were seeded into a 6-well plate 12 h before transfection with indicated combinations of plasmids using the Vigofect reagents (Vigorous). The amounts of plasmids used are 2 μ g for pro-human IL-1 β , 100 ng (HeLa cells) or 50 ng (293T cells) for caspase-1, 100 ng for NLRC4 and 100 ng for NAIP proteins. Twenty-four hours later, LFn-flagellin or LFn-BsaK/PrgJ/CprJ or LFn-CprI together with PA proteins was added into the culture medium at the final concentration of 10 μ g ml⁻¹ for HeLa cells (2 μ g ml⁻¹ for 293T cells) and incubated for another 12 h. Cells were harvested and lysed in a buffer containing 50 mM Tris-HCl (pH 7.6), 150 mM NaCl and 1% Triton X-100. Lysates were resolved onto SDS-PAGE gels followed by anti-IL-1 β immunoblotting analysis. All the reconstitution experiments were performed more than three times and representative results are shown in the figures.

RNAi knockdown. For siRNA knockdown, immortalized BMMs were cultured in 24-well plates at a density of 4 \times 10⁴ per well, and siRNA transfection was performed using the INTERFERin reagent (Polyplus Transfection) by following the manufacturer's instruction. 2 μ l of 20 μ M siRNA (final concentration, 100 nM) and 2 μ l of INTERFERin reagents were used for each well. Sixty hours after transfection, knockdown efficiency and caspase-1 activation were monitored by quantitative real-time PCR (qRT-PCR) and anti-caspase-1 immunoblotting analysis, respectively.

To achieve stable knockdown in immortalized BMMs or U937 cells, shRNAs targeting NAIP5, NAIP2, human NLRC4 or human ASC (listed in Supplementary Table 1) were cloned into a modified lentiviral vector pLKO.1, in which puromycin resistance gene was replaced by GFP coding sequence. pLKO.1-GFP shRNA plasmids were transfected together with two packing plasmids (pCMV-dR8.2 dvpr and pCMV-VSV-G, both from Addgene) into 293T cells. Lentivirus expressing shRNA was collected from the supernatant 48 h after transfection and was used to infect BMMs for another 48 h or U937 cells for 12 h. GFP-positive cells were sorted out by flow cytometry. The pool of sorted cells were either directly used in subsequent functional assays or diluted into 96-well plates to obtain single clones. Knockdown efficiency was examined by qRT-PCR analysis or immunoblotting analysis (for ASC knockdown in U937 cells).

Caspase-1-mediated inflammasome activation assays. To assay caspase-1 activation, culture supernatants of macrophages treated with indicated stimuli were subjected to TCA precipitation and the precipitates were analysed by anti-caspase-1 immunoblotting to detect both pro-Casp1 and processed mature p10 fragment; cell lysates were blotted with Casp1 and actin antibodies to show the level of pro-Casp1 in cell lysates and actin loading, respectively. All caspase-1 activation assays in response to LFn-mediated protein delivery and bacterial infection were repeated at least three times and the representative results are shown in the figures. Mature IL-1 β released into the culture supernatants was measured by using the IL-1 β ELISA kit (Neobioscience Technology Company). Pyroptotic cell death was measured by the lactate dehydrogenase (LDH) assay using CytoTox 96 Non-Radioactive Cytotoxicity Assay kit (Promega). Cell viability was determined by the CellTiter-Glo Luminescent Cell Viability Assay (Promega).

qRT-PCR analysis. For qRT-PCR analysis, total RNA was extracted by TRIzol (Invitrogen) and digested with DNase I (Invitrogen). One microgram of total RNA was reverse-transcribed into cDNA using M-MLV reverse transcriptase (Promega). qRT-PCR analysis was performed using the SYBR Premix Ex Taq (TaKaRa) on Applied Biosystems 7500 Fast Real-Time PCR System. Primers used for qRT-PCR analysis are listed in Supplementary Table 2. The mRNA level of targeted genes was normalized to that of *Gapdh* for mouse BMMs or to that of actin for U937 cells.

Bacterial manipulation and macrophage infection. *L. pneumophila* strains were cultured on buffered charcoal yeast extract agar supplemented with 0.1 mg ml⁻¹ thymidine (BCYET). For infection, bacteria were scraped, diluted in sterile water and added to cells. EPEC strains (E2348/69) were grown overnight in 2 \times YT (tryptone, 16.0 g l⁻¹, yeast extract, 10.0 g l⁻¹, NaCl, 5.0 g l⁻¹) medium without shaking, and then diluted 1:40 in DMEM medium for 4 h to induce the expression of type III secretion system before infection. For *S. typhimurium* infection, overnight 2 \times YT culture was diluted 1:100 and grown for 3 h to induce SPI-1 expression. *B. thailandensis* E264 was obtained from ATCC and cultured as described³¹. Wild-type *C. violaceum* strain (ATCC 12472) was provided by N. Okada and cultured as previously described²⁵. To infect U937 cells, the indicated *C. violaceum* strain cultured overnight at 37 °C in LB broth under conditions of vigorous shaking was diluted 1:100 in fresh LB broth, and further grown for 3 h to obtain an optical density at A₆₀₀ of 2.0 to 2.5. The bacteria were diluted in serum-free RPMI-1640 medium to achieve a multiplicity of infection (MOI) of 10. All infection experiments were performed with a centrifugation of 1,000g for 10 min at 22 °C.

The flagellin-deficient *L. pneumophila* strain (Lp02 Δ fliA) was generated by standard homologous recombination using the suicide plasmid pSR47 s. Deletion of genes encoding the type III rod protein in EPEC (Δ escI) and *S. typhimurium* (Δ prgJ) strains was achieved by using the suicide vector pCVD442 as described previously³⁶. For gene deletion in *B. thailandensis*, a modified suicide vector pDM4-*pheS* expressing a mutant phenylalanine synthetase (PheS) for counter-selection³⁷ was constructed first. Briefly, the *sacB* gene in the commonly used suicide vector pDM4 (provided by D. Milton and L. Gong) was replaced with a 1.1-kb PS12-*pheS* fragment (PS12, the promoter of the *B. pseudomallei* *rpsL* gene) amplified from pBBR1MCS-Km-*pheS* (provided by T. T. Hoang). A PCR fragment containing flanking sequences of the target gene was then cloned into pDM4-*pheS*. The resulting targeting vector was transferred into *B. thailandensis* through *E. coli* SM10 (λ pir)-mediated conjugational mating. The transconjugants were selected in LB agar medium containing chloramphenicol (50 μ g ml⁻¹) and streptomycin (100 μ g ml⁻¹). The integrants were further screened for markerless in-frame deletion by growth on M9 agar plates supplemented with 20 mM glucose and 0.1% p-chlorophenylalanine. All the mutants were verified by PCR and DNA sequencing. Both flagellin genes in *B. thailandensis* E264, *fliC* (open reading frame (ORF), BTH_I3196) and *fliC2* (ORF, BTH_II0151), were deleted to obtain the flagellin-deficient strain. For gene deletion in *C. violaceum*, the original pDM4-*SacB* suicide vector was used. Briefly, a PCR fragment containing flanking sequence of the targeted gene was cloned into pDM4-*SacB*. The resulting targeting vector was transferred into *C. violaceum* through *E. coli* SM10 (λ pir)-mediated conjugational mating. The transconjugants were selected in LB agar medium containing chloramphenicol (17 μ g ml⁻¹) and nalidixic acid (25 μ g ml⁻¹). The integrants were further screened for markerless in-frame deletion by growth on LB agar plates containing 16% sucrose without NaCl. Detailed information for all deletion strains are listed in Supplementary Table 3. All the mutants were verified by PCR and DNA sequencing.

To examine the role of flagellin in stimulating caspase-1 activation during mouse macrophage infection, wild-type, type III secretion-deficient Δ escN (CVD452,

provided by M. Donnenberg) and flagellin-deficient Δ fliC (AGT01, provided by J. A. Girón) strains of EPEC E2348/69 were used to infect immortalized BMMs at a MOI of 10 for 2 h. Wild-type, type III-deficient Δ bipB and flagellin-deficient Δ fliC/*fliC2* strains of *B. thailandensis* were used to infect J774 mouse macrophages at a MOI of 10 for 2 h. To assay the physiological function of NAIP5 in detecting bacterial flagellin, control or *Naip5* stable knockdown immortalized BMMs were infected with *S. typhimurium* strain (wild type, ATCC 14028 s or Δ fliC/ Δ fliB mutant, fliC::Tn10 fliB5001::Mud-Cm; both strains were provided by E. A. Miao) for 15 min, or *L. pneumophila* (Lp02 or Lp02 Δ fliA) for 40 min at a MOI of 50. To assay the function of NAIP2 in detecting the type III rod protein during infection, control or *Naip2* stable knockdown immortalized BMMs were infected with *S. typhimurium* (Δ fliC/ Δ fliB or Δ fliC/ Δ fliB/ Δ prgJ) for 30 min or EPEC E2348/69 strain (Δ fliC or Δ fliC Δ escI) for 2 h at a MOI of 50. Supernatants and cell lysates of infected macrophages were collected and subjected to caspase-1 activation assays described above.

30. Yao, Q. *et al.* A bacterial type III effector family uses the papain-like hydrolytic activity to arrest the host cell cycle. *Proc. Natl Acad. Sci. USA* **106**, 3716–3721 (2009).
31. Cui, J. *et al.* Glutamine deamidation and dysfunction of ubiquitin/NEDD8 induced by a bacterial effector family. *Science* **329**, 1215–1218 (2010).
32. Li, P. *et al.* Mice deficient in IL-1 β -converting enzyme are defective in production of mature IL-1 β and resistant to endotoxic shock. *Cell* **80**, 401–411 (1995).
33. Mariathasan, S. *et al.* Differential activation of the inflammasome by caspase-1 adaptors ASC and Ipaf. *Nature* **430**, 213–218 (2004).
34. Boyden, E. D. & Dietrich, W. F. Nalp1b controls mouse macrophage susceptibility to anthrax lethal toxin. *Nature Genet.* **38**, 240–244 (2006).
35. Vojtek, A. B. & Cooper, J. A. Rho family members: activators of MAP kinase cascades. *Cell* **82**, 527–529 (1995).
36. Dong, N., Liu, L. & Shao, F. A bacterial effector targets host DH-PH domain RhoGEFs and antagonizes macrophage phagocytosis. *EMBO J.* **29**, 1363–1376 (2010).
37. Barrett, A. R. *et al.* Genetic tools for allelic replacement in *Burkholderia* species. *Appl. Environ. Microbiol.* **74**, 4498–4508 (2008).

SUMO1-dependent modulation of SERCA2a in heart failure

Changwon Kho^{1*}, Ahyoung Lee^{1*}, Dongtak Jeong¹, Jae Gyun Oh², Antoine H. Chaanine¹, Eddy Kizana³, Woo Jin Park² & Roger J. Hajjar¹

The calcium-transporting ATPase ATP2A2, also known as SERCA2a, is a critical ATPase responsible for Ca^{2+} re-uptake during excitation-contraction coupling. Impaired Ca^{2+} uptake resulting from decreased expression and reduced activity of SERCA2a is a hallmark of heart failure¹. Accordingly, restoration of SERCA2a expression by gene transfer has proved to be effective in improving cardiac function in heart-failure patients², as well as in animal models³. The small ubiquitin-related modifier (SUMO) can be conjugated to lysine residues of target proteins⁴, and is involved in many cellular processes⁵. Here we show that SERCA2a is SUMOylated at lysines 480 and 585 and that this SUMOylation is essential for preserving SERCA2a ATPase activity and stability in mouse and human cells. The levels of SUMO1 and the SUMOylation of SERCA2a itself were greatly reduced in failing hearts. SUMO1 restitution by adeno-associated-virus-mediated gene delivery maintained the protein abundance of SERCA2a and markedly improved cardiac function in mice with heart failure. This effect was comparable to SERCA2A gene delivery. Moreover, SUMO1 overexpression in isolated cardiomyocytes augmented contractility and accelerated Ca^{2+} decay. Transgene-mediated SUMO1 overexpression rescued cardiac dysfunction induced by pressure overload concomitantly with increased SERCA2a function. By contrast, downregulation of SUMO1 using small hairpin RNA (shRNA) accelerated pressure-overload-induced deterioration of cardiac function and was accompanied by decreased SERCA2a function. However, knockdown of SERCA2a resulted in severe contractile dysfunction both *in vitro* and *in vivo*, which was not rescued by overexpression of SUMO1. Taken together, our data show that SUMOylation is a critical post-translational modification that regulates SERCA2a function, and provide a platform for the design of novel therapeutic strategies for heart failure.

It was previously reported that SERCA2a activity could be modulated by post-translational modifications (PTMs) such as glutathiolation⁶ and nitration⁷. We identified SUMO1 as a SERCA2a interactor in a proteomic screen (Supplementary Fig. 1a, b). We confirmed that SERCA2a binds *in vivo* to SUMO1 and to ubiquitin-conjugating enzyme E2I (UBE2I, also known as UBC9) (Supplementary Fig. 1c–e).

We then examined whether SERCA2a is indeed SUMOylated in human hearts. The SERCA2a band was present along with other slowly migrating SERCA2a bands (Fig. 1a, top panel), which represent the SUMOylated protein. The level of SUMOylated SERCA2a was significantly reduced in failing hearts compared to normal hearts. SUMOylation of SERCA2a was specific for SUMO1 (but not for SUMO2 or 3) (Supplementary Fig. 1f). Along with the reduction in SERCA2a, which is consistent with previous reports, the level of SUMO1 was also significantly reduced in failing hearts. However, the levels of UBC9 and SENP1, critical SUMOylating and de-SUMOylating enzymes, were unaltered in failing hearts (Fig. 1a, bottom panel). These

data indicate that the reduction in SERCA2a SUMOylation might be primarily due to the reduced level of SUMO1 and not to alterations in the level or activity of SUMOylating or de-SUMOylating factors. In additional experiments, we observed that the levels of both SUMO1 and SERCA2a were significantly reduced in a murine model of heart failure induced by pressure overload (Fig. 1b), and in a porcine model of heart failure induced by volume overload (Fig. 1c).

To characterize the potential role of SUMO1 during heart failure, we restored SUMO1 expression in a murine model of transverse aortic constriction (TAC)-induced heart failure. We delivered cardiotropic recombinant adeno-associated viruses, serotype 9 (rAAV9) when mice developed heart failure (Fig. 1d). Gene transfer of rAAV9-SUMO1 resulted in a dose-dependent increase in SUMO1 in the myocardium one month after gene transfer (Supplementary Fig. 2). After induction of heart failure and before gene delivery (two months after TAC), cardiac function was markedly decreased in the mice. The TAC mice received tail-vein injection of either rAAV9-GFP (encoding green fluorescent protein, used as a negative control), rAAV9-SERCA2a (positive control) or rAAV9-SUMO1, and were followed for two more months. Cardiac function assessment by echocardiography revealed substantial improvements after rAAV9-SUMO1 gene delivery (at a total of four months after TAC), despite constant pressure overload, compared with severe deterioration in the rAAV9-GFP control (Fig. 1e, f and Supplementary Table 1). The mice injected with rAAV9-SUMO1 recovered cardiac function to the same degree as the mice injected with rAAV9-SERCA2a. Furthermore, haemodynamic analyses showed substantially improved left-ventricular function in the SUMO1-treated mice, and stabilization of the ratio of heart-weight to body-weight (Supplementary Fig. 3 and Supplementary Table 2). The survival rate in the rAAV9-SUMO1-injected mice was also significantly extended compared with rAAV9-GFP-injected mice with heart failure (Fig. 1g). Consistent with the improvement in cardiac performance in rAAV9-SUMO1-injected mice, protein levels of SERCA2a were significantly increased in rAAV9-SUMO1-injected mice compared with rAAV9-GFP-injected animals, indicating that SUMO1 restoration maintained SERCA2a levels during heart failure (Fig. 1h). Thus, we provide the first evidence, to our knowledge, that cardiac restoration of SUMO1 by rAAV9-mediated gene transfer enables long-term improvement of cardiac function. These findings provide further evidence that manipulating PTMs of SERCA2a may result in a functional benefit.

SUMOylation usually occurs in a highly conserved recognition motif⁸. We found two putative SUMO conjugation sites in human SERCA2a, lysines 480 and 585. These residues are perfectly conserved in mouse, rat, pig and human SERCA2a (Supplementary Fig. 4). We generated three SERCA2a variants with mutations in these lysines. Whereas K480R and K585R were SUMOylated at levels indistinguishable from wild-type SERCA2a, the K480R/K585R mutant was completely un-SUMOylated

¹Cardiovascular Research Center, Mount Sinai School of Medicine, 1 Gustave L. Levy Place, Box 1030, New York, New York 10029, USA. ²College of Life Sciences, Gwangju Institute of Science and Technology, 261 Cheomdan-gwagi-ro, Buk-gu, Gwangju, 500-712, South Korea. ³Westmead Millennium Institute, Sydney Medical School and Department of Cardiology, Westmead Hospital, Westmead NSW 2145, Australia.

*These authors contributed equally to this work.

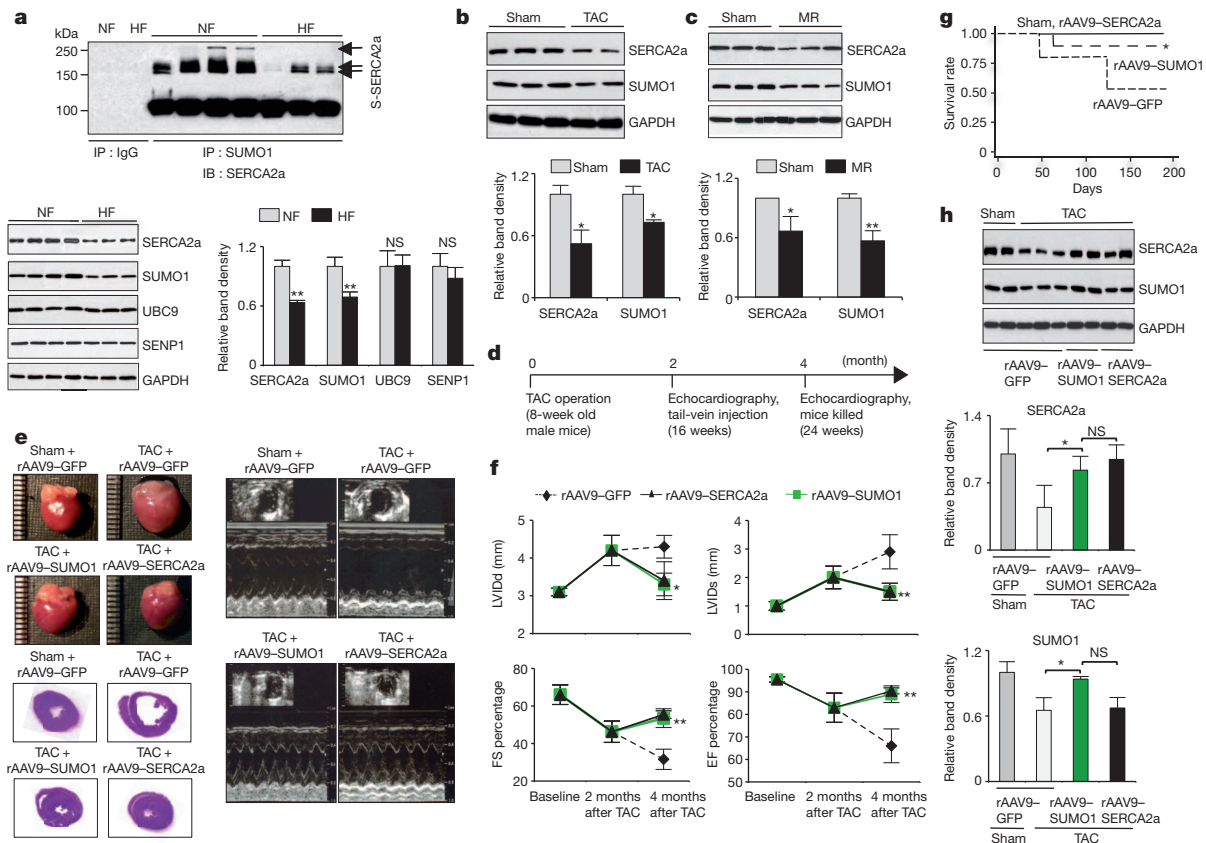


Figure 1 | Endogenous SUMO1 protein levels are decreased in both experimental animal and human heart failure. **a**, SERCA2a SUMOylation in human cardiac tissues ($n = 4$ per heart failure (HF), $n = 3$ per non-failing control (NF)). Representative immunoblots and protein quantification results are shown ($n = 5$ in each group). S-SERCA2a, SUMOylated SERCA2a; GAPDH, glyceraldehyde 3-phosphate dehydrogenase; NS, not significant. **b**, **c**, Representative immunoblots for SUMO1 and SERCA2a, and quantification of protein levels, in models of heart failure. The experimental models are TAC-induced heart failure in mice (**b**; $n = 5$ TAC, $n = 8$ sham-operated control hearts), and porcine heart failure induced by mitral valve regurgitation (MR) (**c**; $n = 3$ MR, $n = 5$ sham-operated). **d**, Protocol for SUMO1 gene therapy in mouse model of heart failure. **e**, Representative gross images of whole hearts (top left panel), haematoxylin and eosin staining results (bottom left panel) and two-dimensional guided M-mode images of the left ventricle (right panel), from rAAV9-SUMO1-injected mice, or negative

(Fig. 2a). Because the SUMOylated lysine residues, K480 and K585, reside in the nucleotide-binding domain that binds ATP, we predicted that SUMOylation might affect the ATPase activity of SERCA2a. As expected, the K480R/K585R mutant showed a significantly decreased ATPase activity compared to the wild type. SUMO1 co-expression significantly increased the ATPase activity of the wild type but did not affect the ATPase activity of K480R/K585R (Fig. 2b). We further tested whether SUMOylation affects the ATP-binding affinity of SERCA2a. Co-expression of SUMO1 markedly increased the ATP-binding affinity of the wild type. By contrast, K480R/K585R had considerably less ATP-binding affinity, which was not affected by SUMO1 co-expression (Fig. 2c). These data indicate that SUMOylation increases the ATPase activity of SERCA2a at least partially by enhancing its ATP-binding affinity.

Because both SERCA2a and SUMO1 levels were reduced in heart failure, we studied whether SUMOylation affects SERCA2a stability. The estimated half-life of the wild type was 4.9 days, which increased to 5.9 days when SUMO1 was co-expressed. The estimated half-life of the K480R/K585R mutant was significantly reduced compared to wild type (Fig. 2d). In endoplasmic-reticulum-associated protein degradation, misfolded proteins are usually degraded by 26S proteasomes after polyubiquitination⁹. To determine whether the instability of the

controls injected with rAAV9-SERCA2a or rAAV9-GFP, subjected to TAC. **f**, Effect of rAAV9-SUMO1 treatment on echocardiography indices of left-ventricle function measured by fractional shortening (FS) and ejection fraction (EF), and left-ventricle end-diastolic diameter measured by internal diameters in end-diastole (LVIDd) and end-systole (LVIDs). TAC rAAV9-GFP, $n = 10$; TAC rAAV9-SUMO1, $n = 14$; TAC rAAV9-SERCA2a, $n = 12$. **g**, Survival of animals after rAAV9-mediated SUMO1 restoration. The Kaplan-Meier method was used to analyse animal lifespan after infection with different viruses. rAAV9-GFP, $n = 12$; rAAV9-SUMO1 or rAAV9-SERCA2a (5×10^{10} vg mouse⁻¹), $n = 24$ in each group. Of 50 TAC-operated animals, 34 mice survived and were further divided into groups that received rAAV9-GFP ($n = 9$), rAAV9-SUMO1 ($n = 13$) or rAAV9-SERCA2a ($n = 12$). **h**, Immunoblot results from a representative experiment and protein quantification ($n = 5$ per group). All data represent the mean \pm s.d. *, $P < 0.05$ versus the respective control, as determined by Student's *t*-test.

K480R/K585R mutant involves ubiquitin-dependent degradation, we performed ubiquitination assays of SERCA2a in HEK-293 cells. SUMO1 overexpression decreased levels of polyubiquitin-conjugated wild-type SERCA2a, whereas it did not alter polyubiquitination of the K480R/K585R mutant (Supplementary Fig. 5). These data support our hypothesis that SUMOylation may compete with ubiquitination on SERCA2a, and indicate that SUMOylation increases the stability of SERCA2a by preventing 26S-proteasomal degradation of SERCA2a.

To examine the physiological function of SUMO1, we isolated cardiomyocytes from adult mice after TAC-induced heart failure or sham operation, and then infected the cells with adenovirus expressing either β -galactosidase (Ad- β -gal) or SUMO1 (Ad-SUMO1). When infected with Ad-SUMO1, cardiomyocytes from sham-operated mice showed significantly enhanced contractility compared to Ad- β -gal-infected cardiomyocytes. A more prominent enhancement in contractility was observed when failing cardiomyocytes were infected with Ad-SUMO1. Ad-SUMO1-infected cardiomyocytes showed increased Ca^{2+} amplitude and faster Ca^{2+} decay compared to the Ad- β -gal-infected controls. The overall inotropic effect of SUMO1 overexpression was comparable to that induced by SERCA2a overexpression (Fig. 3a). It is possible that SUMO1 overexpression

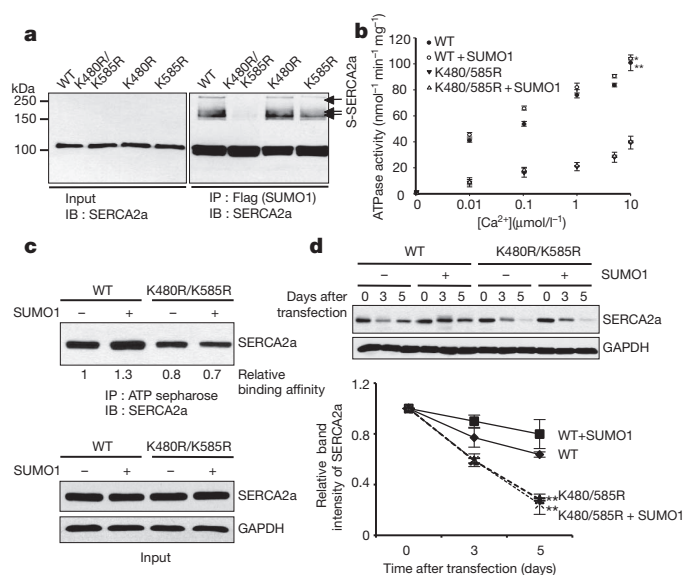


Figure 2 | SUMO1 is conjugated to lysines 480 and 585 of SERCA2a and is required for SERCA2a function. **a**, SUMOylation of SERCA2a *in vivo*. HEK-293 cells were co-transfected with plasmids expressing Flag-tagged SUMO1 and Myc-tagged UBC9 and wild-type or mutant SERCA2a. SUMOylated forms of SERCA2a were detected by immunoblot (IB) analysis using anti-SERCA2a antibody. IP, immunoprecipitation. **b**, Ca²⁺-dependent ATPase activity of wild-type SERCA2a and the K480R/K585R mutant in the presence and absence of additional SUMO1. The data represent three independent experiments, each performed in duplicate. **c**, ATP binding capacity of wild-type and K480R/K585R SERCA2a. Lysates from HEK-293 cells transfected with the indicated plasmids were affinity-precipitated with ATP-sepharose and subsequently subjected to immunoblot analysis with an anti-SERCA2a antibody. **d**, Effects of SUMO1 overexpression on the stability of wild-type and K480R/K585R SERCA2a mutant protein in HEK-293 cells. HEK-293 cells were transfected with wild-type or K480R/K585R SERCA2a expression plasmids, together with empty or SUMO1-expression plasmids. Cells were treated with cycloheximide 48 h after transfection to inhibit protein synthesis. Immunoblot analysis was performed with the indicated antibodies at different time points. Quantification data represent the ratio relative to day 0 (n = 3). All data represent the mean ± s.d. *, P < 0.05; **, P < 0.001 versus the respective control, as determined using Student's *t*-test.

enhanced cardiomyocyte contractility at least partly by increasing the enzymatic activity and stability of SERCA2a.

Given our positive *in vitro* and gene-delivery results, we then proceeded to define the physiological consequences of SUMO1 overexpression *in vivo*. We generated cardiac-specific, Cre/loxP-conditional *Sumo1*-transgenic mice (Supplementary Fig. 6a, b). Along with increased SUMO1 levels, the levels of both SERCA2a itself and its SUMOylation were significantly induced by tamoxifen administration (Supplementary Fig. 6c, d). No cardiac dysfunction was apparent in these transgenic animals (Supplementary Table 3). Wild-type and transgenic mice were subjected to TAC operations and heart failure developed in two months, characterized by a decrease of less than 50% in fractional shortening. One month after tamoxifen was administered (three months in total after TAC), cardiac function was examined. At this point, wild-type mice showed severe failing phenotypes with left ventricular dilation and a low ejection fraction. However, tamoxifen-induced overexpression of SUMO1 markedly reversed these failing phenotypes, resulting in less left ventricular dilation and improved fractional shortening and ejection fraction (Fig. 3b, c and Supplementary Table 4). Haemodynamic analyses also showed improved left ventricular function in the transgenic mice, along with stabilization of heart-weight to body-weight ratio (Supplementary Fig. 7 and Supplementary Table 5). The recovery from cardiac dysfunction induced by SUMO1 overexpression was also manifested by the increased survival rate of transgenic animals under prolonged pressure overload

(Fig. 3d). All transgenic animals survived 100 days after tamoxifen administration, whereas only seven out of fifteen wild-type animals survived at this time point. These findings are consistent with the results from the rAAV9 gene-delivery study.

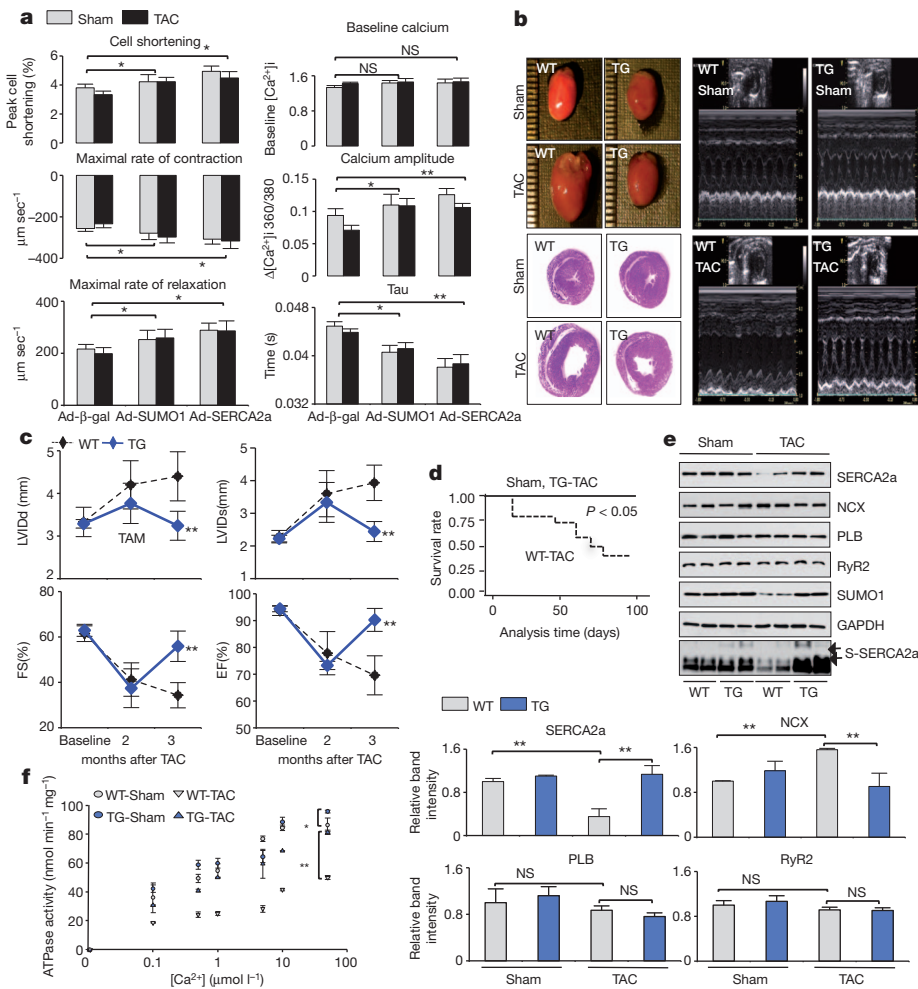
Next, we monitored the expression levels of key regulatory proteins involved in Ca²⁺ homeostasis. After TAC, a notable reduction in the level of SERCA2a was observed, and levels of sodium/calcium exchanger 1 (SLC8A1, also known as NCX1) increased compared to sham-operated controls (Fig. 3e). A decrease in SERCA2a function is coupled with an increase in NCX1 function in failing hearts¹⁰. The TAC-induced changes in the levels of SERCA2a and NCX1 did not occur in transgenic animals.

In wild-type animals, TAC resulted in a significant reduction of SERCA2a ATPase activity, but this was significantly lessened in the transgenic animals (Fig. 3f). Taken together, these data indicate that SUMO1 overexpression ameliorates cardiac dysfunction induced by pressure overload.

To evaluate the effects of downregulation of SUMO1 in hearts, we generated rAAV9 viruses that express SUMO1-directed shRNA (rAAV9-shSumo1), or a scrambled sequence (rAAV9-SC). We confirmed the efficiency and specificity of the shSumo1 in mice (Supplementary Fig. 8a, b). Six weeks after tail-vein injection of viruses, cardiac function was evaluated. Hearts from rAAV9-shSumo1-injected mice showed left ventricular dilation and functional deterioration compared to hearts from rAAV9-SC-injected mice (Fig. 4a, b and Supplementary Tables 6 and 7). Injection of an increased dose of rAAV9-shSumo1 resulted in more severe cardiac dysfunction (Supplementary Fig. 8c, d). Cardiac dysfunction induced by rAAV9-shSumo1 manifested as sudden death of the mice. All mice receiving 1×10^{11} viral genomes (vg) of rAAV9-shSumo1 died within 3 weeks. The death rates of mice that received 3×10^{10} vg and 5×10^{10} vg of rAAV9-shSumo1 were slightly higher than the rates in control mice that received rAAV9-SC (none of these control mice died) (Fig. 4c). SERCA2a protein levels were decreased by approximately 40% in rAAV9-shSumo1-injected hearts, and NCX1 levels were slightly elevated, although this was not statistically significant. As expected, SUMOylation of SERCA2a was also notably blunted (Fig. 4d and Supplementary Fig. 8e). These results support the premise that SUMO1 is important for the stability and function of SERCA2a. Downregulation of SUMO1 also significantly suppressed the ATPase activity of SERCA2a (Fig. 4e), and this was further impaired when a higher dose of rAAV9-shSumo1 was injected (Supplementary Fig. 8f). These data show that SUMO1 is an essential regulator of SERCA2a function in the heart.

We next tested the SUMOylation status of some important cardiac transcriptional factors such as GATA binding protein 4 (GATA4)¹¹ and serum response factor (SRF)¹². These factors are involved in cardiac pathogenesis and have been reported by other groups to be regulated by SUMOylation^{11,12}. The SUMOylation status and protein expression level of GATA4 were not altered by SUMO1 gene delivery, whereas SUMOylation of SRF was increased (Supplementary Fig. 9). Increased SRF SUMOylation may contribute to the activation of certain genes, such as SERCA2a and genes encoding contractile proteins, during heart failure¹³. Several other transcriptional factors, such as SP1 and SP3, and calcium-regulatory signalling pathways, are associated with SERCA2a gene expression^{14,15}. Notably, the cleaved form of SRF was detected in failing hearts and this was decreased by SUMO1 and SERCA2a gene delivery. Recent studies have shown that caspase-3-mediated SRF cleavage occurs in heart failure¹⁶ and SUMO1 overexpression blocks cell-death signalling induced by both tumour necrosis factor and FAS¹⁷. Our findings and other recent studies indicate that SUMO1 restoration increases levels of SUMOylated SRF and may alter apoptotic signalling in heart failure.

To determine the critical link between SUMO1 and SERCA2a in the setting of cardiac dysfunction, we knocked down SERCA2a using a lentiviral system carrying a shRNA directed towards *Serca2a* (shSerca2a). In isolated cardiomyocytes from rat and mouse, shSerca2a markedly



decreased contractile parameters and these were not improved by overexpression of SUMO1 (Supplementary Figs 10a, b and 11a). *In vivo*, rat ventricular function was significantly decreased after gene transfer of sh*Serca2a* but no improvements were observed after co-infection with Ad-*Sumo1* (Supplementary Fig. 10c–f and Supplementary Table 8). In mice, antecedent overexpression of SUMO1 did not rescue ventricular function after lentiviral gene transfer of sh*Serca2a* (Supplementary Fig. 11b–e and Supplementary Table 9). Taken together, these four experiments show that SERCA2a is critical in modulating SUMO1's beneficial effects in the setting of heart failure. However we cannot exclude that knockdown of SERCA2a may have effects on other proteins that would not allow SUMO1 to rescue contractile function.

We have shown that SERCA2a is SUMOylated at two lysine residues. We found that the levels of both SERCA2a itself and SERCA2a SUMOylation were significantly reduced in failing hearts. We have provided compelling evidence that the reduction of SERCA2a SUMOylation is a direct result of reduced SUMO1 levels in failing hearts. This reduction in SUMOylation correlated with reduced ATPase activity and decreased SERCA2a stability. Moreover, restoration of SUMO1 reversed contractile dysfunction in failing hearts. Our findings are summarized in Fig. 4f.

It is possible that SUMOylation may induce a conformational change¹⁸ in SERCA2a or may provide an additional interface for ATP binding, leading to increased ATPase activity. It is also possible that SUMOylation may reciprocally and competitively affect other PTMs of SERCA2a and, in particular, may alter acetylation¹⁹. Notably, acetylation of SERCA2a was recently identified in a large-scale analysis of the human acetylome in cancer cell lines²⁰. We also found that SERCA2a is indeed acetylated and that this acetylation is

more prominent in failing hearts and can be reversed by SIRT1 deacetylase (Kho *et al.*, unpublished data). Further elucidation of the reciprocal or competitive relationship between SUMOylation and acetylation of SERCA2a is currently underway. There are precedents for this type of SUMOylation-mediated inhibition of protein degradation. For example, SUMOylation of axin, a negative regulator of Wnt signalling, prevents ubiquitination and thus induces a prolonged axin half-life²¹. Similarly, SUMOylation of the RNA helicases p68 and p72 increases their stability by reducing ubiquitin-proteasome-mediated protein degradation²². Detailed biochemical analyses are underway to identify the ubiquitination sites in SERCA2a.

Our data show that the amount of SUMO1 is reduced in failing hearts from various species, indicating that the level of cellular SUMO1 must be precisely maintained and controlled for proper cardiomyocyte function. Our finding that restoration of SUMO1 reversed TAC-induced heart failure indicates that a reduction in SUMO1 levels is a direct cause of contractile dysfunction. Our study was based on strong background data that has clearly established impaired SERCA2a as a key molecular abnormality in heart failure. The therapeutic effects of SUMO1, with near-complete recovery of contractile function, are notable, considering that expression of only a single gene is altered. In contrast to the reduction of SUMO1 levels in failing hearts, the levels of the SUMOylating and de-SUMOylating enzymes UBC9 and SENP1 were unaltered, and did not change when shSUMO1 was administered or when SUMO1 levels were restored (data not shown). Therefore, the specificity and capacity of SUMOylation itself is unlikely to be altered in failing hearts. Instead, the supply of SUMO1 itself seems to have the largest impact on SUMOylation. In this regard, it is notable that depletion of cellular ubiquitin is sufficient to cause neuronal dysfunction and death²³.

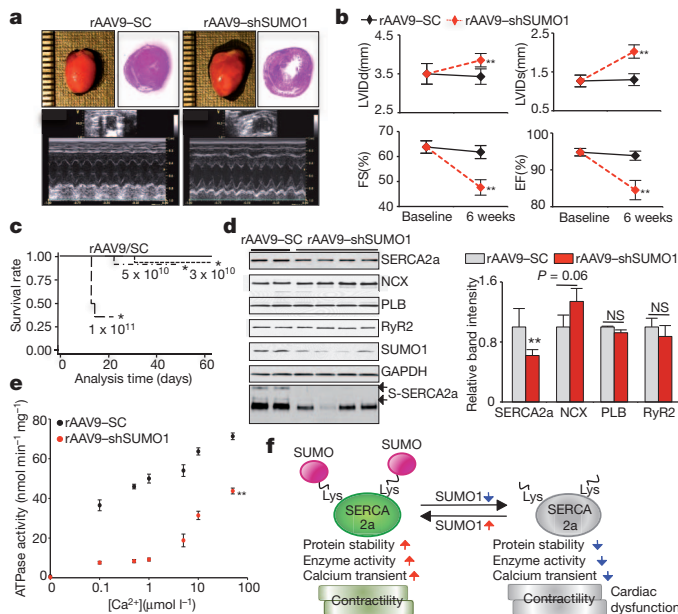


Figure 4 | Reduction of SUMO1 levels accelerates cardiac dysfunction.

a, Heart morphology, cardiac function and heart dimensions were examined in mice 6 weeks after injection of 5×10^{10} vg of rAAVs. Representative gross images of the hearts (top left panel), haematoxylin and eosin staining results (top right panel) and M-mode imaging data (bottom panel) are shown. **b**, Echocardiographic measurements, as in Fig. 1f, for rAAV9-shSC or rAAV9-shSUMO1-injected mice (5×10^{10} vg mouse $^{-1}$, $n = 14$ per group). Kaplan–Meier curves are shown for different doses of rAAV9-shSUMO1 ($n = 14$ per group) or rAAV9-shSC (5×10^{10} vg mouse $^{-1}$, $n = 24$). Solid line represents rAAV9-shSC-injected animals and dotted lines represent rAAV9-shSUMO1-injected animals. **d**, Representative immunoblot analysis for the indicated cardiac proteins in samples collected 6 weeks after tail-vein injection of rAAV9-shSC (5×10^{10} vg mouse $^{-1}$, $n = 4$) or rAAV9-shSUMO1 (5×10^{10} vg mouse $^{-1}$, $n = 7$) (left panel). Quantification of GAPDH-corrected immunoblot signals (right panel). **e**, Ca^{2+} -dependence of ATPase activity of SERCA2a. ATPase activity was examined in preparations from shSC-injected and shSUMO1-injected hearts (5×10^{10} vg mouse $^{-1}$, $n = 3$ per group). All data represent mean \pm s.d. *, $P < 0.05$; **, $P < 0.001$ versus the respective control, as determined using Student's t -test. **f**, A working model for the regulation of SERCA2a function by SUMOylation. Under basal conditions, SUMOylation enhances the stability of SERCA2a and its Ca^{2+} pump function to regulate cardiac contractility. However, increased levels of unSUMOylated SERCA2a due to low SUMO1 protein pools trigger impaired SERCA2a activity and induce cardiac dysfunction under pathophysiological conditions.

In this study we demonstrate a novel regulatory mechanism whereby SUMOylation affects SERCA2a activity and the overall contractile properties of the heart. This highlights the importance of post-translational mechanisms in experimental and human heart failure. In addition, the beneficial effects of SUMO1 on cardiac contractility and survival indicate that targeting SUMO1 may provide a novel therapeutic strategy for the treatment of heart failure.

METHODS SUMMARY

Animals. All mice and rats were housed and treated in accordance with guidelines from the NIH and institutional animal care and use committees, and the protocols used were approved by the Mount Sinai School of Medicine animal care and use committee.

Human heart samples. Left ventricular samples were obtained from explanted human hearts obtained at the time of cardiac transplantation. Non-failing hearts (which were used as controls) were obtained from donors who died from neurological diseases or motor-vehicle accidents, and who had normal cardiac function by echocardiography. The five donors, three males and two females, had a median

age of 62. The heart-failure patients, three males and two females, had a median age of 60 and their average ejection fraction was $20 \pm 3\%$.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 18 February; accepted 1 August 2011.

Published online 7 September 2011.

- Meyer, M. *et al.* Alterations of sarcoplasmic reticulum proteins in failing human dilated cardiomyopathy. *Circulation* **92**, 778–784 (1995).
- Jessup, M. *et al.* Calcium upregulation by percutaneous administration of gene therapy in cardiac disease (CUPID): a phase 2 trial of intracoronary gene therapy of sarcoplasmic reticulum Ca^{2+} -ATPase in patients with advanced heart failure. *Circulation* **124**, 304–313 (2011).
- Kawase, Y. *et al.* Reversal of cardiac dysfunction after long-term expression of SERCA2a by gene transfer in a pre-clinical model of heart failure. *J. Am. Coll. Cardiol.* **51**, 1112–1119 (2008).
- Johnson, E. S. Protein modification by SUMO. *Annu. Rev. Biochem.* **73**, 355–382 (2004).
- Geiss-Friedlander, R. & Melchior, F. Concepts in sumoylation: a decade on. *Nature Rev. Mol. Cell Biol.* **8**, 947–956 (2007).
- Adachi, T. *et al.* S-Glutathiolation by peroxynitrite activates SERCA during arterial relaxation by nitric oxide. *Nature Med.* **10**, 1200–1207 (2004).
- Knyushko, T. V., Sharov, V. S., Williams, T. D., Schoneich, C. & Bigelow, D. J. 3-Nitrotyrosine modification of SERCA2a in the aging heart: a distinct signature of the cellular redox environment. *Biochemistry* **44**, 13071–13081 (2005).
- Sampson, D. A., Wang, M. & Matunis, M. J. The small ubiquitin-like modifier-1 (SUMO-1) consensus sequence mediates Ubc9 binding and is essential for SUMO-1 modification. *J. Biol. Chem.* **276**, 21664–21669 (2001).
- Plemper, R. K. & Wolf, D. H. Retrograde protein translocation: ERADication of secretory proteins in health and disease. *Trends Biochem. Sci.* **24**, 266–270 (1999).
- Schillinger, W., Fiolet, J. W., Schlotthauer, K. & Hasenfuss, G. Relevance of Na^{+} - Ca^{2+} exchange in heart failure. *Cardiovasc. Res.* **57**, 921–933 (2003).
- Wang, J., Feng, X. H. & Schwartz, R. J. SUMO-1 modification activated GATA4-dependent cardiogenic gene activity. *J. Biol. Chem.* **279**, 49091–49098 (2004).
- Matsuzaki, K. *et al.* Serum response factor is modulated by the SUMO-1 conjugation system. *Biochem. Biophys. Res. Commun.* **306**, 32–38 (2003).
- Wang, J. & Schwartz, R. J. Sumoylation and regulation of cardiac gene expression. *Circ. Res.* **107**, 19–29 (2010).
- Brady, M. *et al.* Sp1 and Sp3 transcription factors are required for trans-activation of the human SERCA2 promoter in cardiomyocytes. *Cardiovasc. Res.* **60**, 347–354 (2003).
- Vlasblom, R. *et al.* Contractile arrest reveals calcium-dependent stimulation of SERCA2a mRNA expression in cultured ventricular cardiomyocytes. *Cardiovasc. Res.* **63**, 537–544 (2004).
- Drewett, V. *et al.* Serum response factor cleavage by caspases 3 and 7 linked to apoptosis in human BJAB cells. *J. Biol. Chem.* **276**, 33444–33451 (2001).
- Okura, T. *et al.* Protection against Fas/APO-1- and tumor necrosis factor-mediated cell death by a novel protein, sentrin. *J. Immunol.* **157**, 4277–4281 (1996).
- Baba, D. *et al.* Crystal structure of thymine DNA glycosylase conjugated to SUMO-1. *Nature* **435**, 979–982 (2005).
- Van Rechem, C. *et al.* Differential regulation of HIC1 target genes by CtBP and NuRD, via an acetylation/SUMOylation switch, in quiescent versus proliferating cells. *Mol. Cell. Biol.* **30**, 4045–4059 (2010).
- Choudhary, C. *et al.* Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* **325**, 834–840 (2009).
- Kim, M. J., Chia, I. V. & Costantini, F. SUMOylation target sites at the C terminus protect axin from ubiquitination and confer protein stability. *FASEB J.* **22**, 3785–3794 (2008).
- Mooney, S. M., Grande, J. P., Salisbury, J. L. & Janknecht, R. Sumoylation of p68 and p72 RNA helicases affects protein stability and transactivation potential. *Biochemistry* **49**, 1–10 (2010).
- Ryu, K. Y., Garza, J. C., Lu, X. Y., Barsh, G. S. & Kopito, R. R. Hypothalamic neurodegeneration and adult-onset obesity in mice lacking the *Ubb* polyubiquitin gene. *Proc. Natl Acad. Sci. USA* **105**, 4016–4021 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work is supported by NIH RO1 HL083156, HL080498, HL093183 and P20HL100396 (R.J.H.). W.J.P. is funded by Global Research Laboratory Program (M6-0605-00-0001) of the Korean Ministry of Science and Technology.

Author Contributions C.K., A.L. and R.J.H. conceived the project and its design. C.K., A.L., D.J., J.G.O. and A.H.C. performed experiments and data analysis. E.K. aided in experimental design. C.K., A.L., W.J.P. and R.J.H. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.J.H. (roger.hajjar@mssm.edu).

METHODS

In vivo SUMOylation assay. To analyse SUMOylation in cells, lipofectamine 2000 was used to transfect HEK-293 cells with plasmids encoding wild-type or SUMOylation-site mutants of SERCA2a, along with Flag-tagged SUMO1 and Myc-tagged UBC9. The cells were lysed by sonication in ice-cold lysis buffer (50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 0.1% Triton X-100, 10 mM EDTA, complete protease inhibitor (1 tablet per 10 ml, Roche) and protein phosphatase inhibitor cocktail (Sigma)) containing 20 mM *N*-ethylmaleimide. Lysates were cleared by centrifugation at 30,000g for 20 min. Cell lysates were then immunoprecipitated overnight at 4 °C with a Flag-specific affinity matrix gel (Sigma), after which the immunoprecipitates were washed in cold lysis buffer. The immunocomplexes were then resolved by SDS polyacrylamide gel electrophoresis (SDS-PAGE) and immunoblot analysis was performed with a SERCA2a-specific antibody.

Fresh tissue extracts were prepared in lysis buffer for *in vivo* SUMOylation assays. Hearts from each experimental and control group were frozen in liquid nitrogen. The frozen tissues were crushed and homogenized in lysis buffer as described above using the MP homogenate system (FastPrep homogenizer). The insoluble portion was removed by centrifugation at 30,000g for 20 min. The extracts were incubated with anti-SUMO1 agarose resin with agitation overnight. The SUMO-conjugated forms were detected by immunoblot analysis with specific primary antibodies.

SERCA2a activity assay. SERCA2a activity was determined using an assay based on pyruvate/NADH-coupled reactions, as previously described²⁴. The activity of the Ca^{2+} -ATPase was calculated as follows: $\Delta\text{absorbance}/6.22 \times \text{protein} \times \text{time}$ (in nmol ATP per mg protein \times min). All assays were performed in triplicate.

Generation of conditional *Sumo1*-transgenic mice. The αMHC -flox-mouse SUMO1 transgene was subcloned into the pML2G vector (gift from Y. Wang), which encodes GFP cDNA between two loxP sites. The construct was microinjected into fertilized eggs from B6C3 mice, and transgenic integration was confirmed by PCR (mouse genetics shared research facility, Mount Sinai School of Medicine).

Antibodies. Antibodies against the following proteins were used for immunoblotting: SERCA2a (21st Century Biochemicals), GAPDH (Sigma), Myc (Sigma), Flag (Sigma), SUMO1 (Cell Signaling), SUMO1 agarose resin (Santa Cruz Biotech), SUMO2/3 (MBL), UBC9 (Boston Biochem), SENP1 (Pierce), PLN (Badrilla), RyR2 (Badrilla), NCX1 (Chemicon), GATA4 (Millipore), SRF (Millipore), anti-rabbit-HRP and anti-mouse-HRP secondary antibodies (Sigma).

Primers for mutagenesis. The following primer sets were used for mutagenesis: K480R_sense, 5'-TCA GTC ATT AAA CAG CTG ATG AGA AAG GAA TTC ACT CTA GAG-3'; K480R_antisense, 5'-CTC TAG AGT GAA TTC CTT TCT CAT CAG CTG TTT AAT GAC TGA-3'; K585R_sense, 5'-GAG GAC TCT GCC AAC TTT ATT AGA TAT GAG ACC AAT CTG ACC-3'; K585R_antisense, 5'-GGT CAG ATT GGT CTC ATA TCT AAT AAA GTT GGC AGA GTC CTC-3'.

Proteomic analysis. Solubilized immunocomplexes were combined with a rehydration buffer (9 M urea, 2 M thiourea, 4 M CHAPS buffer, 16 mM dithiothreitol, 2% w/v pharmalyte 3-11 and trace amounts of bromophenol blue) to a final volume of 340 μl , and then rehydrated for 16 h. After rehydration, strips were focused at 60 kVh at 20 °C (IPGphor III, GE Healthcare). When isoelectric focusing was complete, the strips were equilibrated, separated on 12.5% SDS-PAGE gels and visualized by silver staining using the Plus One silver staining kit (GE Healthcare), with minor modifications to ensure compatibility with subsequent mass spectrometry analysis. Stained two-dimensional gel images were scanned (ImageScanner II, GE Healthcare) and analysed by PDquest 8.1 software system (Bio-Rad). Spots were excised from the gels by spot picker and prepared for mass spectrometry as previously described²⁵.

Mass spectrometry analysis. Identification of gel spots was accomplished by ESI/LC/MS/MS system (Thermo-Scientific). Full scan spectra were recorded in positive mode over the mass range 350–2000 Da. MS/MS data were automatically acquired on the two most intense precursor ions in each full scan spectrum, and were interpreted using Xcalibur 2.0 and Bioworks 3.2 software.

Contractility measurement. The mechanical properties of isolated ventricular cardiomyocytes were assessed using a video-based edge detection system (IonOptix) as previously described²⁶.

Adenoviruses. Adenoviruses encoding SERCA2a and SUMO1 were generated using the pAdEasy XL adenoviral vector system (Stratagene) according to the manufacturer's protocols.

rAAV vector production and purification. rAAV9-SERCA2a, rAAV9-shSumo1 and rAAV9-SC were produced using the two-plasmids protocol as previously described²⁷ with the following modifications: HEK-293T cells (ATCC) were grown in triple flasks for 24 h (DMEM, 10% fetal bovine serum) before adding the calcium phosphate precipitate. After 72 h, the virus was purified from benzonase-treated crude cell lysates over an iodixanol density gradient (Optiprep, Greiner Bio-One Inc.), followed by heparin-agarose type I affinity

chromatography (Sigma). Finally, viruses were concentrated and formulated into lactated Ringer's solution (Baxter Healthcare Corporation) using Vivaspin 20 centrifugal concentrators 50K MWCO (Vivascience Inc.), and stored at -80°C .

Lentivirus. The lentiviral plasmids encoding shRNAs for *Serca2a* were previously described by us²⁸. The *Serca2a* shRNA construct is the following sequence; the underlined segment indicates the loop: 5'-GATCCGACTTACTAGTTAGAA TTTGGCTAAGAGCAAATTTCTAACTAGTAAGTCTTTTGGAAATTAAT-3'. We used third-generation lentiviral systems to generate the vectors²⁹.

Transverse aortic constriction. Mice underwent TAC using a supraclavicular construction model as previously described²⁶. TAC or sham surgery was performed in 8-week-old male mice (body weight 18–22 g). Mice were anesthetized with intraperitoneal ketamine and placed on a ventilator. A longitudinal cut of 2–3 mm was made in the proximal portion of the sternum, allowing visualization of the aortic arch. The transverse aortic arch was ligated between the innominate and left common carotid arteries with an overlying 27-gauge needle. The needle was immediately removed leaving a discrete region of constriction. The sham group underwent a similar procedure without ligation.

Gene transfer *in vivo*. Adult Sprague-Dawley rats (body weight 240–260 g) underwent gene transfer via aortic cross-clamping as previously described³⁰. After dissection of the aorta and pulmonary artery, lentiviral sh*Serca2a* with or without adenoviral SUMO1 were injected into the left ventricular cavity through a 22G catheter while the aorta and pulmonary artery were cross-clamped for 50 s. In sham-operated animals, normal saline was injected into the left ventricular cavity while the aorta and pulmonary artery were cross-clamped for 50 s. After cross-clamping was released, the chest was closed.

Echocardiography. Transthoracic echocardiography was performed using a Vivid 7000 (GE Healthcare) equipped with a H13L transducer (14 MHz). Two-dimensional and M-mode images were obtained in the short-axis view. The heart rate, left ventricular end-diastolic internal diameter and left ventricular end-systolic internal diameter were measured in at least three repeated cardiac cycles. The ejection fraction and fractional shortening were then calculated.

Haemodynamics. Haemodynamic measurements are performed using a 1.2 Fr pressure-volume conductance catheter (Sciense). Pressure-volume loop analysis was performed as previously described³¹. Mice were injected intraperitoneally with urethane (1 g kg^{-1}), etomidate (10 mg kg^{-1}) and morphine (1 mg kg^{-1}) and mechanically ventilated with $7\text{ }\mu\text{l g}^{-1}$ stroke volume at 125 respirations min^{-1} . The chest was opened to expose the heart for an apical stab approach. To determine absolute ventricular volumes via admittance technology, myocardial and blood conductance were obtained before pressure-volume catheter placement in the left ventricle³². The inferior vena cava was transiently occluded to reduce ventricular pre-load to obtain load-independent pressure-volume relationships. Haemodynamic measurements were acquired and analysed using IOX software (EMKatech).

Histological analysis. Hearts were prepared in Tissue-Tek OCT compound (Sakura Finetechnical) and sectioned into 6- μm slices (Microm HM560 Cryo-star, Thermo Scientific). The sections were stained with haematoxylin and eosin.

Statistical analysis. Statistical analyses were performed using Student's *t*-test. Significant differences are indicated *, $P < 0.05$ or **, $P < 0.001$. Data in the figures represent mean \pm s.d.

24. Hajjar, R. J., Schmidt, U., Kang, J. X., Matsui, T. & Rosenzweig, A. Adenoviral gene transfer of phospholamban in isolated rat cardiomyocytes. Rescue effects by concomitant gene transfer of sarcoplasmic reticulum Ca^{2+} -ATPase. *Circ. Res.* **81**, 145–153 (1997).
25. Lee, A. Y. *et al.* Identification of the degradome of Isp-1, a major intracellular serine protease of *Bacillus subtilis*, by two-dimensional gel electrophoresis and matrix-assisted laser desorption/ionization-time of flight analysis. *Proteomics* **4**, 3437–3445 (2004).
26. Jeong, D. *et al.* PICOT inhibits cardiac hypertrophy and enhances ventricular function and cardiomyocyte contractility. *Circ. Res.* **99**, 307–314 (2006).
27. Zolotukhin, S. *et al.* Recombinant adeno-associated virus purification using novel methods improves infectious titer and yield. *Gene Ther.* **6**, 973–985 (1999).
28. Kizana, E., Cingolani, E. & Marban, E. Non-cell-autonomous effects of vector-expressed regulatory RNAs in mammalian heart cells. *Gene Ther.* **16**, 1163–1168 (2009).
29. Tiscornia, G., Singer, O., Ikawa, M. & Verma, I. M. A general method for gene knockdown in mice by using lentiviral vectors expressing small interfering RNA. *Proc. Natl Acad. Sci. USA* **100**, 1844–1848 (2003).
30. Hajjar, R. J. *et al.* Modulation of ventricular function through gene transfer *in vivo*. *Proc. Natl Acad. Sci. USA* **95**, 5251–5256 (1998).
31. Pacher, P., Nagayama, T., Mukhopadhyay, P., Batkai, S. & Kass, D. A. Measurement of cardiac function using pressure-volume conductance catheter technique in mice and rats. *Nature Protocols* **3**, 1422–1434 (2008).
32. Porterfield, J. E. *et al.* Dynamic correction for parallel conductance, G_p , and gain factor, α , in invasive murine left ventricular volume measurements. *J. Appl. Physiol.* **107**, 1693–1703 (2009).

The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes

Tian-Peng Gu^{1*}, Fan Guo^{1*}, Hui Yang^{2*}, Hai-Ping Wu^{1†}, Gui-Fang Xu¹, Wei Liu¹, Zhi-Guo Xie¹, Linyu Shi², Xinyi He³, Seung-gi Jin⁴, Khursheed Iqbal⁵, Yujiang Geno Shi⁶, Zixin Deng³, Pirooska E. Szabó⁵, Gerd P. Pfeifer⁴, Jinsong Li² & Guo-Liang Xu¹

Sperm and eggs carry distinctive epigenetic modifications that are adjusted by reprogramming after fertilization¹. The paternal genome in a zygote undergoes active DNA demethylation before the first mitosis^{2,3}. The biological significance and mechanisms of this paternal epigenome remodelling have remained unclear⁴. Here we report that, within mouse zygotes, oxidation of 5-methylcytosine (5mC) occurs on the paternal genome, changing 5mC into 5-hydroxymethylcytosine (5hmC). Furthermore, we demonstrate that the dioxygenase Tet3 (ref. 5) is enriched specifically in the male pronucleus. In Tet3-deficient zygotes from conditional knockout mice, paternal-genome conversion of 5mC into 5hmC fails to occur and the level of 5mC remains constant. Deficiency of Tet3 also impedes the demethylation process of the paternal *Oct4* and *Nanog* genes and delays the subsequent activation of a paternally derived *Oct4* transgene in early embryos. Female mice depleted of Tet3 in the germ line show severely reduced fecundity and their heterozygous mutant offspring lacking maternal Tet3 suffer an increased incidence of developmental failure. Oocytes lacking Tet3 also seem to have a reduced ability to reprogram the injected nuclei from somatic cells. Therefore, Tet3-mediated DNA hydroxylation is involved in epigenetic reprogramming of the zygotic paternal DNA following natural fertilization and may also contribute to somatic cell nuclear reprogramming during animal cloning.

To investigate whether loss of DNA methylation in the male pronucleus coincides with oxidation of 5mC to 5hmC, a recently reported type of modification of mammalian DNA^{5,6}, we performed immunostaining of mouse zygotes using an antibody specifically recognizing 5hmC (Supplementary Fig. 1). We found that the 5hmC signal increased markedly in the paternal pronucleus around the pronuclear stage PN3 when the paternal pronucleus became larger than the maternal pronucleus (Fig. 1a, b). By contrast, the 5mC signal became markedly weaker in the male pronucleus from the PN3 stage whereas there was no clear change in the female pronucleus (Supplementary Fig. 2), as reported previously². The inverse correlation between the 5hmC and 5mC signals in the two parental genomes seemed to persist beyond the zygotic stage (Supplementary Fig. 3). Therefore, 5mC oxidation in the male pronucleus coincides with the loss of methylation in the early mouse embryo. A similar observation has recently been reported in two independent studies^{7,8}.

Next, we examined the expression of Tet enzymes that can catalyse the oxidation of 5mC in DNA⁹. The *Tet3* mRNA was specifically detected in oocytes and zygotes (Supplementary Fig. 4). At the zygotic stage, the Tet3 protein was concentrated in the male pronucleus, but localized to the cytoplasm at other pre-implantation stages (Fig. 1c and Supplementary Fig. 5). The unique expression pattern of Tet3 suggests its possible role in modifying the zygotic paternal genome.

To study the biological function of Tet3 in mouse, we generated a conditional knockout allele abolishing its catalytic activity (Supplementary Fig. 6a). Because homozygous mutation led to neonatal lethality, we achieved germ-line-specific deletion of Tet3 from primordial germ cells (PGCs) in [*Tet3*^{f/f}, TNAP-Cre] conditional knockout (CKO) mice.

Female CKO mice were normal in growth and morphology. Although they displayed much reduced fecundity (see below), they gave birth to heterozygous offspring when crossed with wild-type males (Supplementary Fig. 6b). The deletion of Tet3 in oocytes and zygotes was confirmed by immunostaining and PCR with reverse transcription assays (Fig. 2a and Supplementary Fig. 6c). Strikingly, no 5hmC signal could be detected and the 5mC signal intensity did not decline in the late male pronuclei of zygotes collected from the CKO females mated with wild-type males (Fig. 2b and Supplementary Fig. 7). In contrast, deletion of *Tet3* from the male germ cells did not seem to affect the change in 5hmC and 5mC. Therefore, the loss of 5mC in the paternal genome in developing zygotes is caused by its conversion to 5hmC and the maternal Tet3 is required for this conversion.

We then assessed the role of Tet3 in demethylation of specific sequences in the male pronucleus. *Line1* transposons are known to be actively demethylated in zygotes^{10,11}. Comparison of the methylation level of male pronuclear DNA from Tet3-deficient zygotes at the PN3–4 stages with that of wild-type zygotes showed that the process of active DNA demethylation was impeded by Tet3 deletion (Fig. 2c, upper panel). This finding also indicates that 5hmC serves as an intermediate between 5mC and unmethylated C, although bisulphite analysis cannot distinguish 5hmC from 5mC^{12,13}. To confirm this, we assayed for 5mC and 5hmC on *Line1* elements in the paternal DNA by MeDIP and hMeDIP (methylated and hydroxymethylated DNA immunoprecipitation). 5hmC was indeed present at *Line1* sequences in wild-type zygotic male pronuclei at a significantly higher level compared to sperm, whereas the 5mC level was markedly lower than that of sperm (Supplementary Fig. 8). In the analysis of male pronuclear DNA from zygotes lacking Tet3, enrichment of 5hmC-containing *Line1* elements was significantly decreased whereas enrichment of 5mC-containing elements was comparable with sperm. These results strengthen the conclusion that 5mC oxidation does occur at *Line1* sequences in male pronuclei.

Embryonic stem cell marker genes, such as *Oct4* and *Nanog*, are methylated during male germ cell development and their demethylation occurs in the early embryo^{14–16}. In normal zygotes, the *Oct4* gene of male pronuclei had undergone substantial demethylation by the PN3–4 pronuclear stages, but this process was markedly hampered in Tet3-null zygotes (Fig. 2c and Supplementary Fig. 9). Moreover, Tet3 deletion almost completely blocked demethylation at two other

¹Group of DNA Metabolism, The State Key Laboratory of Molecular Biology, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China. ²The State Key Laboratory of Cell Biology, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China.

³The State Key Laboratory of Microbial Metabolism, School of Life Science and Biotechnology, Shanghai Jiaotong University, Shanghai 200030, China. ⁴Department of Cancer Biology, Beckman Research Institute of the City of Hope, Duarte, California 91010, USA. ⁵Department of Molecular and Cellular Biology, Beckman Research Institute of the City of Hope, Duarte, California 91010, USA. ⁶Division of Endocrinology, Diabetes, and Hypertension, Department of Medicine and BCMP, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA. [†]Present address: Novartis Institutes for BioMedical Research Co., Shanghai 201203, China.

*These authors contributed equally to this work.

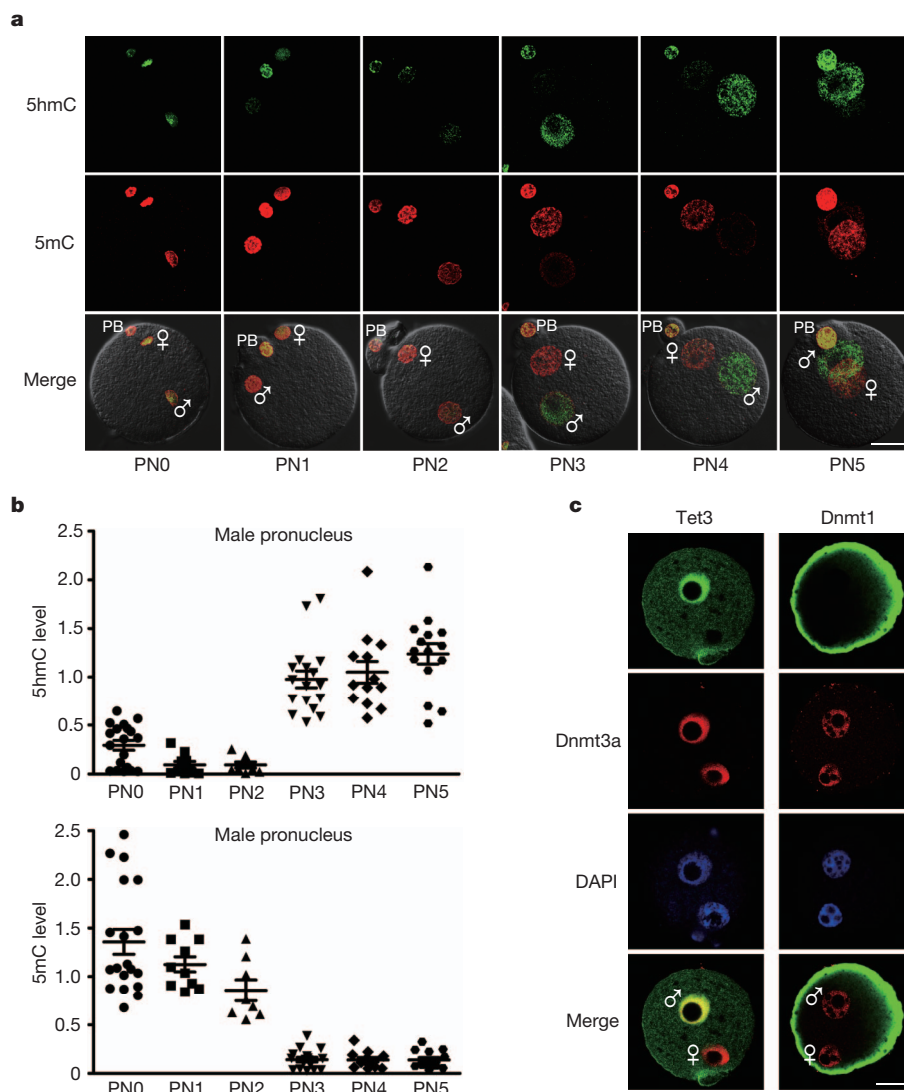


Figure 1 | Specific oxidation of methylcytosine and Tet3 distribution in the zygotic male pronucleus. **a**, Immunofluorescent images of 5hmC (green) and 5mC (red) staining, and overlaid phase contrast images. The pronuclear (PN) stages are indicated. Male and female symbols indicate male and female pronucleus, respectively. PB, polar body. Scale bar, 25 μ m. **b**, Quantification of the relative levels of 5hmC and 5mC in male pronuclei in zygotes. Each data point is based on the level of the 5hmC or 5mC signal relative to the DAPI

staining intensity of the same pronucleus. Error bars indicate s.e.m. Number of zygotes analysed for each stage: PN0, 19; PN1, 10; PN2, 8; PN3, 17; PN4, 17; PN5, 15. **c**, Preferential staining of Tet3 protein in the male pronucleus. DNA was stained with DAPI. Control staining shows Dnmt1 in the cortical cytoplasm and Dnmt3a (red) in both pronuclei. The nucleolus had no staining signal.

paternally methylated genes, *Nanog* and *Lemdl* (ref. 17), which retained hypermethylation similar to that observed in sperm (Supplementary Fig. 9). To assess the significance of paternal demethylation on gene expression, Tet3-null oocytes from the CKO females were fertilized by intracytoplasmic injection (ICSI) of wild-type sperm carrying the enhanced green fluorescent protein (EGFP) reporter gene under the control of the *Oct4* promoter, and expression of EGFP was monitored in cultured embryos. Compared with wild-type embryos, the mutant embryos derived from oocytes lacking Tet3 showed significantly weaker EGFP expression at the 8-cell and morula stages (Fig. 2d and Supplementary Fig. 10). Based on these global and sequence-specific analyses of 5mC and 5hmC, along with the reporter gene assay, we conclude that Tet3-mediated 5mC oxidation contributes to the demethylation in the zygotic paternal genome and gene activation in the early embryo.

We next investigated whether removing Tet3 from oocytes might compromise embryonic development. We first confirmed that early deletion of Tet3 from the PGC stage did not affect epigenetic reprogramming in the embryonic germ cells, oocyte development,

maturation and fertilization (Supplementary Figs 11–13 and Supplementary Table 1). Male germ cell development and sperm DNA methylation were not affected either (Supplementary Fig. 14). However, fecundity of the female CKO mice was significantly lower in terms of the frequency of successful pregnancy per mating and the litter size (Fig. 3a and Supplementary Table 2).

Deletion of maternal Tet3 did not seem to affect the pre-implantation development as heterozygous zygotes collected from CKO females mated with wild-type males developed to blastocysts *in vitro* normally (Supplementary Table 3). We then examined the effect of maternal Tet3 deletion on prenatal development by transplantation of 2-cell embryos into oviducts of pseudo-pregnant females. Whereas the transferred embryos lacking maternal Tet3 implanted normally, they showed a much reduced rate of full-term development (Supplementary Table 4). Dissection of transferred [*Tet3* *Mat*[−]/*Pat*⁺] mutant embryos in the pregnant wild-type foster mothers revealed a high frequency of degeneration and the appearance of morphological abnormalities, starting from midgestation (Fig. 3b and Supplementary Table 4). Moreover, deletion of Tet3 later from growing oocytes

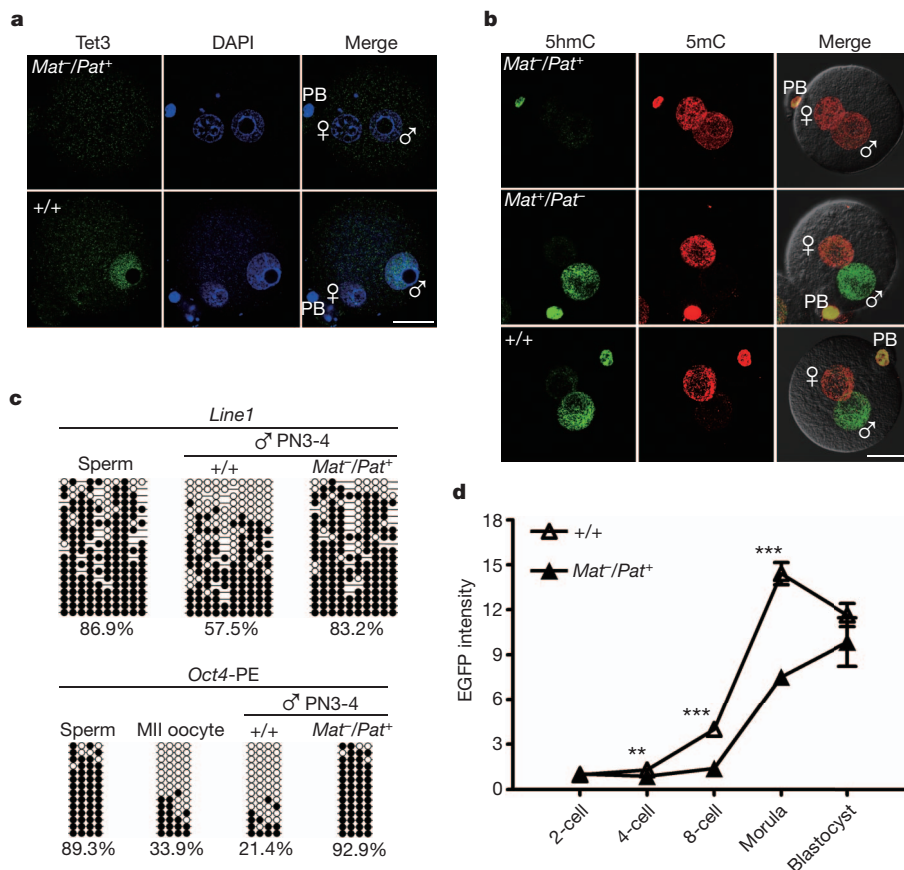


Figure 2 | The role of Tet3 in 5mC oxidation, demethylation of paternal DNA, and activation of the paternal *Oct4* allele. **a**, Loss of the Tet3 protein in heterozygous zygotes (*Mat⁻/Pat⁺*) obtained from CKO females mated with wild-type males. PN3 zygotes were stained with an anti-Tet3 antibody raised against the deleted region. Scale bar, 25 μ m. **b**, 5hmC (green) and 5mC (red) immunostaining in wild-type (*+/+*) and maternally (*Mat⁻/Pat⁺*) or paternally (*Mat⁺/Pat⁻*) Tet3-deficient zygotes at the PN5 stage. **c**, Methylation analysis of *Line1* and *Oct4* (PE, the proximal enhancer region) in male pronuclei isolated from wild-type (*+/+*) and Tet3-deficient (*Mat⁻/Pat⁺*) zygotes. Open and filled circles represent unmethylated and methylated CpG sites, respectively. Percentage of methylated CpGs is indicated. **d**, Paternal *Oct4* activation in wild-type embryos (*+/+*) and embryos lacking maternal Tet3 (*Mat⁻/Pat⁺*). Embryos were derived from ICSI using sperm carrying the *Oct4-EGFP* transgene and the EGFP signal was quantified, relative to the level in a 2-cell blastomere. Number of embryos analysed at each stage, *+/+*: 2-cell, 13; 4-cell, 9; 8-cell, 9; morula, 12; blastocyst, 7. *Mat⁻/Pat⁺*: 2-cell, 11; 4-cell, 9; 8-cell, 8; morula, 8; blastocyst, 5. Error bars indicate s.e.m. ***P* < 0.01 and ****P* < 0.001.

using Zp3-Cre also led to failure in zygotic 5hmC generation, retention of paternal 5mC, impaired demethylation at *Line1* and *Oct4*, and compromised embryonic development (Supplementary Fig. 15). Therefore, lacking maternal Tet3 blocks paternal genome reprogramming and causes markedly increased developmental failure of the embryo.

Somatic cell nuclei injected into eggs undergo profound epigenetic reprogramming, including DNA demethylation. The cytoplasm of germinal vesicle (GV) oocytes¹⁸, metaphase II (MII) oocytes¹⁹, zygotes²⁰ and 2-cell embryos²¹ has been shown to possess reprogramming activity. The existence of the Tet3 protein across these stages indicated that Tet3 might be one of the cytoplasmic factors contributing to the reprogramming activity. We tested therefore whether Tet3-mediated hydroxylation is part of the reprogramming process in somatic cell nuclear transfer (SCNT). Remarkably, following activation of nuclear transfer (NT) oocytes reconstructed after injection of somatic nucleus into oocyte with or without enucleation (intact oocyte), the Tet3 protein originating from the oocyte cytoplasm became concentrated in the pseudo-pronucleus (PPN) formed from the transferred somatic nucleus, but not in the female pronucleus derived from the spindle-chromosome-complex that existed in oocytes (Fig. 4a). Significantly, the PPN formed in the NT embryos from wild-type intact or enucleated oocytes underwent 5mC oxidation whereas this modification did not occur in embryos derived from Tet3-null oocytes (Fig. 4b). Whereas substantial demethylation was detected at the *Oct4* promoter in the PPN of Tet3-proficient NT embryos, the somatic hypermethylation persisted when the oocyte Tet3 was deleted (Fig. 4c). To investigate the role of Tet3-mediated DNA oxidation in the activation of pluripotency genes, we used donor somatic cells carrying the *Oct4-EGFP* transgene. Compared with embryos from wild-type oocytes, the EGFP signal was significantly lower in embryos derived from the Tet3-null oocytes at the

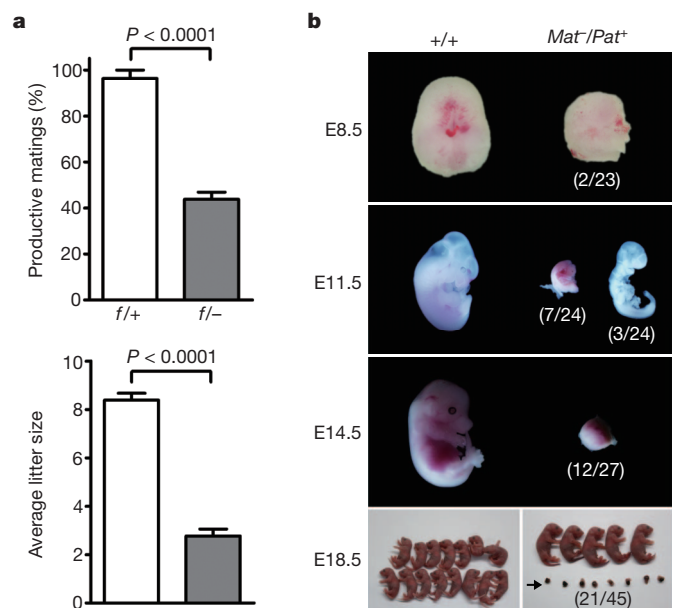


Figure 3 | Maternal Tet3 deficiency compromises embryonic development. **a**, Reduced efficiency of productive mating and litter size in females (*f/-*) with germline Tet3 deficiency. Error bars indicate s.e.m. [*Tet3^{f/+}*, TNAP-Cre] mice, *n* = 9; [*Tet3^{f/-}*, TNAP-Cre] mice, *n* = 7. **b**, Developmental failure among embryos lacking maternal Tet3. Pregnant foster females receiving wild-type (*+/+*) and mutant (*Mat⁻/Pat⁺*) 2-cell embryos transferred were dissected at different stages of gestation. Numbers in brackets indicate proportion of embryonic day 8.5 (E8.5) deciduas and E11.5–18.5 embryos showing a smaller size and morphological abnormalities. The arrow indicates degenerated conceptuses obtained. Detailed data are presented in Supplementary Table 4.

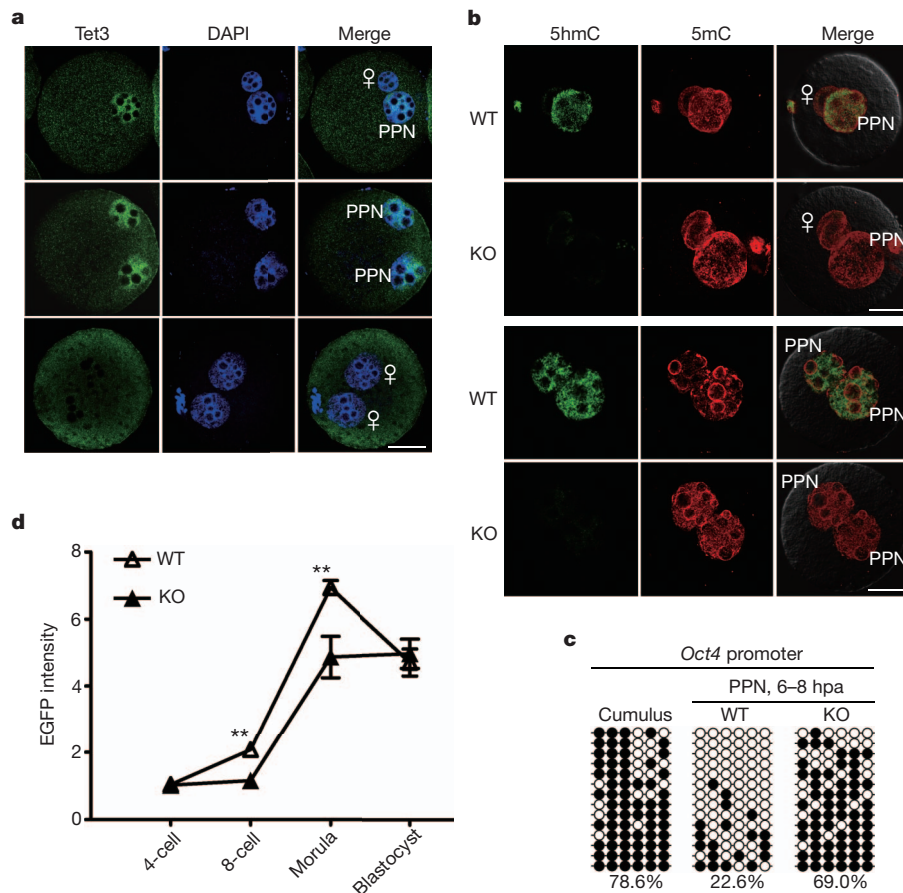


Figure 4 | Tet3 contributes to reprogram the somatic nucleus transferred into oocytes. **a**, Tet3 enrichment in the pseudo-pronucleus (PPN) formed in embryos reconstructed by transplantation of cumulus somatic nuclei into an intact (top row) or enucleated oocyte (middle row). The enrichment did not occur in the female pronucleus in intact oocytes (top row) and in the two pronuclei of parthenogenetic embryos (bottom row). Scale bar, 25 μ m. **b**, Impaired 5mC oxidation in PPN of 1-cell NT embryos derived from injection of cumulus nuclei into intact or enucleated oocytes lacking Tet3. Embryos around 10 h post activation (hpa) were stained with anti-5hmC (green) and

anti-5mC (red) antibodies. **c**, Impaired *Oct4* demethylation in NT embryos derived from Tet3-null oocytes. More than 20 NT embryos derived from enucleated wild-type (WT) or Tet3-null (KO) oocytes were collected 6–8 hpa for DNA methylation analysis. **d**, Weakened activation of the somatic *Oct4-EGFP* reporter in cultured NT embryos from Tet3-null oocytes. The EGFP intensities were relative to the level in a 4-cell blastomere. Error bars indicate s.e.m. Number of embryos analysed for each stage, WT: 4-cell, 11; 8-cell, 13; morula, 9; blastocyst, 7. KO: 4-cell, 8; 8-cell, 7; morula, 6; blastocyst, 7. ** $P < 0.01$.

8-cell and morula stages (Fig. 4d and Supplementary Fig. 16a). Consistently, the *Oct4* mRNA level was relatively low in 8-cell embryos from Tet3-null oocytes (Supplementary Fig. 16b). Therefore, deficiency in oocyte Tet3 could cause weakened or delayed activation of the somatic *Oct4* in NT embryos.

We have obtained substantial evidence that oxidation of 5mC in the paternal genome in fertilized eggs by Tet3 initiates DNA demethylation and facilitates the activation of the paternal copy of early embryonic genes, thus contributing to the establishment of biparental totipotency in the early embryo by counteracting the silencing function of 5mC. Blocking oxidation of 5mC by Tet3 deletion affects paternal gene activation, leading to reduced developmental fitness and fetal survival. It remains to be determined whether the developmental failures could be caused by haploinsufficiency²² for the genes affected in the early embryo.

The involvement of Tet3 in somatic *Oct4* activation indicates that Tet-mediated 5mC oxidation contributes to epigenetic reprogramming of the donor nuclear DNA in SCNT. Reprogramming in SCNT might thus share a common mechanism with paternal genome remodelling in fertilized eggs. Further investigations are needed to reveal the signals regulating Tet3 and the events subsequent to 5hmC formation in both fertilized and cloned embryos.

METHODS SUMMARY

Preparation of anti-5hmC antibody. To prepare anti-5hmC antibody, 5-hydroxymethylcytidine 5'-monophosphate (5hmCMP) hapten was synthesized from cytidine 5'-monophosphate (CMP) using Mita hydroxymethylase²³. The ribonucleoside was then conjugated to bovine serum albumin (BSA) as previously described²⁴ and used to immunize rabbits. The antibody was affinity-purified from antiserum with 5hmCMP-BSA conjugate coupled to agarose beads.

Immunostaining for 5hmC in fertilized oocytes. Immunofluorescence detection of 5hmC in zygotes followed the procedure for 5mC¹¹. Fluorescent images were acquired at 2- μ m Z-axis intervals using a confocal microscope (LEICA TCS SP5 II) and their signal intensity was determined.

Gene targeting. A *Tet3* targeting vector was electroporated into 129Sv ES cells for homologous recombination. The floxed region contains exons 8–9, which includes the coding region for the conserved Fe²⁺-binding motif of dioxygenases. To inactivate *Tet3* in the germ line, we crossed the mice carrying a floxed allele with TNAP-Cre knock-in mice on a 129 genetic background²⁵. The conditional knock-out mice were on C57BL/6J-129Sv genetic background.

Zygote collection and staging. Female mice 4- to 8-week-old were injected with pregnant mare's serum gonadotropin and human chorionic gonadotropin were mated with wild-type males. Zygotes were harvested and the PN stages of individual zygotes were classified according to ref. 11, by taking into account the pronuclear morphology and the presence of 5mC signal.

Observation of *Oct4-EGFP* activation. To examine the role of Tet3 in paternal gene activation, MII oocytes collected from wild-type and Tet3 CKO females were

fertilized by ICSI with sperm from transgenic mice (tg/tg) carrying the *Oct4-EGFP* transgene²⁶ and the resulting embryos were cultured to observe EGFP expression. To examine the role of Tet3 in somatic gene activation in SCNT, we monitored the *Oct4-EGFP* transgene expression from the donor cell DNA in NT embryos, which were derived after injection of a cumulus nucleus into a wild-type or a Tet3-null oocyte without enucleation.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 21 January; accepted 16 August 2011.

Published online 4 September 2011.

- Surani, M. A., Hayashi, K. & Hajkova, P. Genetic and epigenetic regulators of pluripotency. *Cell* **128**, 747–762 (2007).
- Mayer, W., Niveleau, A., Walter, J., Fundele, R. & Haaf, T. Demethylation of the zygotic paternal genome. *Nature* **403**, 501–502 (2000).
- Oswald, J. *et al.* Active demethylation of the paternal genome in the mouse zygote. *Curr. Biol.* **10**, 475–478 (2000).
- Ooi, S. K. & Bestor, T. H. The colorful history of active DNA demethylation. *Cell* **133**, 1145–1148 (2008).
- Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).
- Kriaucinis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929–930 (2009).
- Iqbal, K., Jin, S. G., Pfeifer, G. P. & Szabo, P. E. Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proc. Natl Acad. Sci. USA* **108**, 3642–3647 (2011).
- Wossidlo, M. *et al.* 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nature Commun.* **2**, 241 (2011).
- Wu, S. C. & Zhang, Y. Active DNA demethylation: many roads lead to Rome. *Nature Rev. Mol. Cell Biol.* **11**, 607–620 (2010).
- Jamil, A. Z., Iqbal, K., Fawad Ur, R. & Mirza, K. A. Effect of phacoemulsification on intraocular pressure. *J. Coll. Physicians Surg. Pak.* **21**, 347–350 (2011).
- Wossidlo, M. *et al.* Dynamic link of DNA demethylation, DNA strand breaks and repair in mouse zygotes. *EMBO J.* **29**, 1877–1888 (2010).
- Jin, S. G., Kadam, S. & Pfeifer, G. P. Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. *Nucleic Acids Res.* **38**, e125 (2010).
- Huang, Y. *et al.* The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PLoS ONE* **5**, e8888 (2010).
- Hattori, N. *et al.* Epigenetic control of mouse *Oct-4* gene expression in embryonic stem cells and trophoblast stem cells. *J. Biol. Chem.* **279**, 17063–17069 (2004).
- Imamura, M. *et al.* Transcriptional repression and DNA hypermethylation of a small set of ES cell marker genes in male germline stem cells. *BMC Dev. Biol.* **6**, 34 (2006).
- Farthing, C. R. *et al.* Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes. *PLoS Genet.* **4**, e1000116 (2008).
- Iqbal, K. *et al.* Subcutaneous panniculitis-like T-cell lymphoma in association with sarcoidosis. *Clin. Exp. Dermatol.* **36**, 677–679 (2011).
- Bui, H. T. *et al.* The cytoplasm of mouse germinal vesicle stage oocytes can enhance somatic cell nuclear reprogramming. *Development* **135**, 3935–3945 (2008).
- Yang, H. *et al.* High-efficiency somatic reprogramming induced by intact MII oocytes. *Cell Res.* **20**, 1034–1042 (2010).
- Egli, D., Rosains, J., Birkhoff, G. & Eggan, K. Developmental reprogramming after chromosome transfer into mitotic mouse zygotes. *Nature* **447**, 679–685 (2007).
- Egli, D., Sandler, V. M., Shinohara, M. L., Cantor, H. & Eggan, K. Reprogramming after chromosome transfer into mouse blastomeres. *Curr. Biol.* **19**, 1403–1409 (2009).
- Seidman, J. G. & Seidman, C. Transcription factor haploinsufficiency: when half a loaf is not enough. *J. Clin. Invest.* **109**, 451–455 (2002).
- Li, L. *et al.* The mildiomycin biosynthesis: initial steps for sequential generation of 5-hydroxymethylcytidine 5'-monophosphate and 5-hydroxymethylcytosine in *Streptovorticillium rimofaciens* ZJU5119. *ChemBioChem* **9**, 1286–1294 (2008).
- Erlanger, B. F. & Beiser, S. M. Antibodies specific for ribonucleosides and ribonucleotides and their reaction with DNA. *Proc. Natl Acad. Sci. USA* **52**, 68–74 (1964).
- de Vries, W. N. *et al.* Expression of Cre recombinase in mouse oocytes: a means to study maternal effect genes. *Genesis* **26**, 110–112 (2000).
- Ohbo, K. *et al.* Identification and characterization of stem cells in prepubertal spermatogenesis in mice. *Dev. Biol.* **258**, 209–225 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Walsh and M. Rots for critical reading of the manuscript, J. Walter for discussions, H. Qi for providing cDNA of mouse oocytes, R. Zhang & Q. Cui for *Tet3* cDNA, L. Li for help with 5hmCMP synthesis, Shanghai Research Center for Model Organisms for blastocyst injection, and J. Gao for mouse work. This study was supported by grants from the Ministry of Science and Technology China (2007CB947503 to G.-L.X., 2007CB947101 to J.L., and 2009CB941101 to G.-L.X. and J.L.), National Science Foundation of China (30730059 to G.-L.X. and 30871430 to J.L.), the Chinese Academy of Sciences (XDA01010301 to G.-L.X.; XDA01010403 and KSCX2-YW-R-110 to J.L.) and the NIH (GM078458 to Y.G.S.).

Author Contributions G.-L.X. and J.L. conceived the projects. Y.G.S., H.-P.W. and G.-L.X. contributed to the knockout design. F.G., T.-P.G., H.-P.W., G.-F.X., and W.L. performed the experiments on early embryos. X.H. and Z.D. contributed to the synthesis of the 5hmC hapten. H.Y. and L.S. performed the nuclear transfer and embryo transfer experiments. S.-g.J., K.L., P.E.S., G.P.P. and Z.-G.X. characterized Tet3 expression in PGCs and ovaries. G.-L.X. wrote and G.P.P. revised the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.L. (jsli@sibs.ac.cn) or G.-L.X. (glxu@sibs.ac.cn).

METHODS

Preparation of anti-5hmC and anti-Tet3 antibodies. To prepare 5hmC antibody, 5-hydroxymethylcytidine 5'-monophosphate (5hmCMP) hapten was synthesized from cytidine 5'-monophosphate (CMP) in a reaction containing formaldehyde and tetrahydrofolate catalysed by recombinant MilA hydroxymethylase²³. The ribonucleoside hapten was then conjugated to BSA as previously described²⁴ for immunization of rabbits (Supplementary Fig. 1). Mass spectral analysis confirmed that the 5hmCMP base moiety was unaltered under the conjugation condition. The antibody was affinity-purified from antiserum with the 5hmCMP-BSA conjugate coupled to agarose beads. To ensure specificity, cross-reactivity was removed by incubation with agarose beads crosslinked with CMP-BSA conjugate.

For the detection of Tet3, two rabbit polyclonal antibodies were raised against a C-terminal region (amino acids 1159–1329, GenBank accession number NP_898961) and the targeted region (amino acids 887–962) respectively, affinity-purified and evaluated as described previously²⁷.

Immunostaining for 5hmC and Tet3 in fertilized oocytes. Immunofluorescence detection of 5hmC in zygotes derived from natural matings followed the procedure for 5mC¹¹. The signal was detected by Alexa Fluor-conjugated goat anti-rabbit or anti-mouse IgG (see Supplementary Table 4 for detailed antibody information). Fluorescent images were acquired at 2- μ m Z-axis intervals using a confocal microscope (Leica TCS SP5 II) and their signal intensity was determined using the Leica Application Suite-Advanced Fluorescence software.

The distribution of the Tet3 protein in embryonic cells was determined on embryos fixed with 4% paraformaldehyde (PFA) and permeabilized with 0.5% Triton.

Gene targeting. A targeting vector for *Tet3* was prepared using the recombineering technique²⁸ and electroporated into 129Sv ES cells for selection of targeted clones. The floxed region contains exons 8–9, which code for the region (76 amino acids from EEVLR to NGCTV, GenBank accession number NP_898961) containing the conserved Fe²⁺-binding motif of the catalytic domain. Deletion of the floxed region leads to the loss of 76 amino acids with in-frame fusion between exons 7 and 10. Neomycin-resistant embryonic stem clones were screened by PCR using a pair of primers crossing the shorter right homologous arm. Positive clones were further characterized by Southern blotting to confirm homologous recombination on the left side of the targeted genomic region (refer to Supplementary Fig. 7a). Embryonic stem cells carrying a correctly targeted allele (with *neo*) were injected into blastocysts to generate germline chimaeras. Mice with a floxed allele were obtained by breeding with C57BL/6J mice. The *neo* selection marker was removed in mice by crossing with ACTFLPe mice²⁹. To inactivate *Tet3* in germ cells from the PGC stage onwards, we generated conditional knockout mice by crossing floxed mice with TNAP-Cre knock-in mice. TNAP-Cre is expressed in primordial germ cells from embryonic day 9.5 to late gestation³⁰. To inactivate *Tet3* in female germ cells from the growing oocyte stage, we generated conditional knockout mice by crossing floxed mice with Zp3-Cre transgenic mice which express Cre exclusively in growing oocytes³¹. Mice were genotyped by PCR (primer sequences are presented in Supplementary Table 6). The conditional knockout mice were on a mixed C57BL/6J-129Sv genetic background.

Zygote collection. Wild-type BDF1 (from C57BL/6 \times DBA2 σ) female mice 4- to 8-week-old injected with pregnant mare's serum gonadotropin and human chorionic gonadotropin were mated with BDF1 male. Zygotes were harvested at different time points after human chorionic gonadotropin injection. The PN stage of each individual zygote was classified according to ref. 11, by taking into account the pronuclear morphology and the presence of 5mC signal. We stained the DNA with DAPI (for fixed zygotes) or Hoechst 33258 (for live zygotes), and used a Nomarski differential interference contrast microscope for better observation of zygotic pronuclei, when necessary. Tet3-deficient zygotes were obtained from [*Tet3*^{fl/fl}, TNAP-Cre] female mice crossed with wild-type males.

Fertility test. [*Tet3*^{fl/fl}, TNAP-Cre] female mice at 6–8 weeks of age were housed with 8- to 12-week-old wild-type males of proven fertility. Rate of fertilization was judged by the presence of two pronuclei in zygotes collected from females after pregnant mare's serum gonadotropin and human chorionic gonadotropin treatment and mating with wild-type male mice of proven fertility. Number of completed pregnancies per plug seen (litters per plug) and number of viable pups born per litter (litter size) were calculated from the data included in Supplementary Table 2. [*Tet3*^{fl/+}, TNAP-Cre] female mice were used as a control group for comparison. Student's *t*-test was performed to compare averages in the two different experimental groups and *P* < 0.05 was considered to be significant.

Isolation of primordial germ cells (PGCs). Female CFI (Charles River Laboratories) mice were mated with male OG2 (ref. 32) mice. This transgenic mouse line expresses the enhanced green fluorescent protein (EGFP) from the *Oct4* promoter and thus enables the selective purification of embryonic and fetal germ cells. Embryo parts enriched in PGCs and genital ridges were dissected at 9.5

and 11.5 days post coitum (dpc), respectively. At 9.5 dpc and 11.5 dpc, the sex of the embryo was determined by real-time PCR amplification of two genes in a single reaction. The *Sry* amplicon indicated the presence of Y chromosome and male sex, whereas amplification of the *Snrpn* gene served as a positive control for DNA. Gonads were dissected from male and female fetuses at 13.5, 15.5 and 17.5 dpc. Male and female gonads were distinguished by their distinct morphology at these stages. Gonads were incubated at 37 °C for 15 min in trypsin-EDTA and triturated to achieve a single cell suspension containing germ cells and somatic cells. Dulbecco's Modified Eagle's Medium (DMEM) (Invitrogen) supplemented with 20% FBS was added to inactivate trypsin. Cell suspensions were analysed and sorted on a MoFlo flow cytometer (Beckman Coulter). Data were acquired using 488 nm excitation from an Innova-306 Argon laser (Coherent) at 500 mW. EGFP emission was measured through a 530DF30 filter (Omega Optical).

Quantitative reverse transcription PCR. Poly(A⁺) mRNAs were isolated from zygotes (*n* = 200) and female PGCs (embryonic days 9.5, 11.5, 13.5, 15.5 and 17.5) by using the Dynabeads mRNA DIRECT Micro Kit (Invitrogen). Oligo (dT)25-coupled Dynabeads and mRNA complexes were immediately used for reverse transcription using SuperScript III reverse transcriptase (Invitrogen), according to the manufacturer's instructions. Real-time quantitative PCR reactions were performed at 50 °C for 2 min and 95 °C for 10 min followed by 50 cycles at 95 °C for 15 s and 60 °C for 1 min using TaqMan Gene Expression Master Mix (Applied Biosystems) on an iQ5 real-time PCR cycler (Bio-Rad). PCR was performed with TaqMan MGB primers with 6FAM-based probes (Applied Biosystems) using the following assay ID numbers: *Tet1* (Mm01169088_m1), *Tet2* (Mm01312907_m1), *Tet3* (Mm00805754_m1) and *Stella/Dppa3* (Mm01184198_g1). The cDNA levels of target genes were analysed using a comparative *C_t* method and normalized to the internal standard gene *Gapdh*.

Collection of oocytes and production of parthenogenetic embryos. For collection of GV oocytes, the ovaries were removed from the female mice 42–44 h after pregnant mare's serum gonadotropin injection. Antral follicles were punctured by 30G needles, and cumulus-enclosed GV oocytes were released into HEPES-buffered CZB medium (HCZB) containing 0.2 mM 3-isobutyl-1-methylxanthine (IBMX) to inhibit germinal vesicle breakdown. Cumulus cells were removed by pipetting. For collection of mature oocytes, oviducts were removed from the female mice 13–15 h after human chorionic gonadotropin injection. Cumulus-oocyte complexes were released into HCZB containing 0.1% bovine testicular hyaluronidase (300 USP units per mg; ICN Biomedicals Inc.). MII oocytes were activated for 6 h in activation medium (calcium-free CZB medium containing 10 mM Sr²⁺ and 5 μ g ml⁻¹ cytochalasin B) to generate parthenogenetic embryos, which were cultured in KSOM medium with amino acids and harvested at 8 h post activation.

Intracytoplasmic sperm injection (ICSI). ICSI was performed according to the method of ref. 33 except for being performed at room temperature (about 25 °C). Briefly, sperm were collected from adult mice (Oct-delta PE-GFP #18) carrying *Oct4-EGFP* transgene (tg/tg)²⁶ and the head was separated from the tail by applying pulses to the head-tail junction by means of a Piezo-driven pipette (Piezoelectric actuator; PrimeTech). Only the sperm head was injected into each oocyte. Injected oocytes were cultured in KSOM medium for 96 h to examine their development *in vitro*. Images of resulting embryos were acquired with an IX51 inverted microscope (Olympus) under the same exposure parameters and the EGFP intensity of each embryo was quantified with ImageJ software³⁴.

Embryo transfer and Caesarean section. Fertilized eggs derived from natural matings were cultured in KSOM medium until the 2-cell stage. Two-cell embryos were then transferred into oviducts of surrogate females at day 1 of pseudopregnancy. For strict comparison, eight mutant and control 2-cell embryos (Supplementary Table 4) were transferred into the left and right oviducts of recipients, respectively. Recipient mothers were euthanized at 8.5, 11.5 and 14.5 days of gestation and embryos were dissected. For embryos developed to term, caesarean section was performed on day 19 and living pups were nursed by lactating ICR females.

Isolation of male pronuclei. Male pronuclei, which were distinguished from female pronuclei on the basis of their size and distance from polar bodies were harvested from zygotes of PN3–4 stages by breaking the zona using Piezo drive (Prime Tech) and aspirating using a micromanipulator. At least 40 male pronuclei from control or Tet3-deficient zygotes were collected and subjected to bisulphite sequencing analysis.

Bisulphite sequencing. For DNA methylation analysis in oocytes, pronuclei from zygotes and pseudo-pronuclei from NT embryos with limited numbers, bisulphite conversion was performed in agarose beads as described³⁵. Unbiased amplification for methylated and unmethylated sequences was ensured by testing bisulphite PCR primers using a 1:1 mixture of unmethylated and *in vitro* methylated DNA fragments. The PCR products were cloned into pMD19-T vectors (Takara Inc.) and individual clones were sequenced by BGI Ltd, Shanghai. Bisulphite primer

information is presented in Supplementary Table 6. For the determination of the methylation state of each sequence, the experiment was performed at least twice starting from the isolation of cells, pronuclei and embryos.

DNA immunoprecipitation with anti-5mC and anti-5hmC antibodies. To detect the existence of 5hmC at *Line1* repeats in the paternal genome in mouse zygotes, >100 male pronuclei were harvested from zygotes at PN3–4 stages, digested with proteinase K and RNase A, and the genomic DNA was purified by phenol-chloroform extraction. The genomic DNA was mixed with 250 ng of carrier lambda DNA (*dam*[−], *dcm*[−]) and fragmented by AluI digestion, heat-denatured (10 min, 95 °C), and immunoprecipitated as described previously³⁶ using 1 µg of anti-5hmC or anti-5mC antibodies (Eurogentec, BI-MECY-1000) and 10 µl Dynabeads (coupled with M-280 sheep anti-rabbit IgG for the 5hmC antibody or anti-mouse IgG for the 5mC antibody). qPCR was performed on a BioRad CFX96 Real-Time PCR Detection System for the input and immunoprecipitated DNA. Mouse genomic DNA (10 ng) from Dnmt TKO embryonic stem cells lacking DNA methylation³⁷ and thus containing no 5hmC, was used as negative control. Mouse sperm DNA was used for comparisons.

Nuclear transfer with intact and enucleated oocytes. NT was performed as described³⁸ with modifications¹⁹. Briefly, metaphase II-arrested oocytes were collected from superovulated B6D2F1 or CKO females, and cumulus cells were removed using hyaluronidase. In the standard NT procedure, the oocytes were collected from wild-type and Tet3 CKO females, and enucleated in a droplet of HEPES-CZB medium containing 5 µg ml^{−1} CB using a blunt Piezo-driven pipette. After enucleation, the spindle-free oocytes were washed extensively and maintained in CZB medium up to 2 h before nucleus injection. The cumulus cells collected from superovulated *Oct4-EGFP* transgenic mice were aspirated in and out of the injection pipette to remove the cytoplasmic material and then injected into enucleated oocytes. The reconstructed oocytes were cultured in CZB medium for 1 h and then activated for 5–6 h in activation medium. For NT with intact oocytes, oocytes were activated for 20 min and then directly injected with cumulus cells. The reconstructed oocytes were activated for 5–6 h in activation medium.

Following activation, the reconstructed embryos were cultured in KSOM medium with amino acids at 37 °C under 5% CO₂ in air. Embryo imaging and EGFP quantification followed the same procedure as in the ICSI experiment described above. The EGFP levels were determined from *n* > 6 embryos at each stage.

27. Ge, Y. Z. *et al.* Chromatin targeting of de novo DNA methyltransferases by the PWWP domain. *J. Biol. Chem.* **279**, 25447–25454 (2004).
28. Liu, P., Jenkins, N. A. & Copeland, N. G. A highly efficient recombineering-based method for generating conditional knockout mutations. *Genome Res.* **13**, 476–484 (2003).
29. Rodriguez, C. I. *et al.* High-efficiency deleter mice show that FLPe is an alternative to Cre-loxP. *Nature Genet.* **25**, 139–140 (2000).
30. Lomeli, H., Ramos-Mejia, V., Gertsenstein, M., Lobe, C. G. & Nagy, A. Targeted insertion of Cre recombinase into the TNAP gene: excision in primordial germ cells. *Genesis* **26**, 116–117 (2000).
31. Lewandoski, M., Wassarman, K. M. & Martin, G. R. *Zp3-cre*, a transgenic mouse line for the activation or inactivation of *loxP*-flanked target genes specifically in the female germ line. *Curr. Biol.* **7**, 148–151 (1997).
32. Szabó, P. E., Hubner, K., Scholer, H. & Mann, J. R. Allele-specific expression of imprinted genes in mouse migratory primordial germ cells. *Mech. Dev.* **115**, 157–160 (2002).
33. Kimura, Y. & Yanagimachi, R. Intracytoplasmic sperm injection in the mouse. *Biol. Reprod.* **52**, 709–720 (1995).
34. Abràmoff, M. D., Magalhães, P. J. & Ram, S. J. Image processing with ImageJ. *Biophotonics Int.* **11**, 36–42 (2004).
35. Hajkova, P. *et al.* DNA-methylation analysis by the bisulfite-assisted genomic sequencing method. *Methods Mol. Biol.* **200**, 143–154 (2002).
36. Weber, M. *et al.* Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nature Genet.* **37**, 853–862 (2005).
37. Tsumura, A. *et al.* Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells* **11**, 805–814 (2006).
38. Wakayama, T., Perry, A. C., Zuccotti, M., Johnson, K. R. & Yanagimachi, R. Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei. *Nature* **394**, 369–374 (1998).

Conformational changes in the G protein Gs induced by the β_2 adrenergic receptor

Ka Young Chung^{1*}, Søren G. F. Rasmussen^{1,2*}, Tong Liu^{3*}, Sheng Li³, Brian T. DeVree⁴, Pil Seok Chae^{5†}, Diane Calinski⁴, Brian K. Kobilka¹, Virgil L. Woods Jr³ & Roger K. Sunahara⁴

G protein-coupled receptors represent the largest family of membrane receptors¹ that instigate signalling through nucleotide exchange on heterotrimeric G proteins. Nucleotide exchange, or more precisely, GDP dissociation from the G protein α -subunit, is the key step towards G protein activation and initiation of downstream signalling cascades. Despite a wealth of biochemical and biophysical studies on inactive and active conformations of several heterotrimeric G proteins, the molecular underpinnings of G protein activation remain elusive. To characterize this mechanism, we applied peptide amide hydrogen–deuterium exchange mass spectrometry to probe changes in the structure of the heterotrimeric bovine G protein, Gs (the stimulatory G protein for adenylyl cyclase) on formation of a complex with agonist-bound human β_2 adrenergic receptor (β_2 AR). Here we report structural links between the receptor-binding surface and the nucleotide-binding pocket of Gs that undergo higher levels of hydrogen–deuterium exchange than would be predicted from the crystal structure of the β_2 AR–Gs complex. Together with X-ray crystallographic and electron microscopic data of the β_2 AR–Gs complex (from refs 2, 3), we provide a rationale for a mechanism of nucleotide exchange, whereby the receptor perturbs the structure of the amino-terminal region of the α -subunit of Gs and consequently alters the ‘P-loop’ that binds the β -phosphate in GDP. As with the Ras family of small-molecular-weight G proteins, P-loop stabilization and β -phosphate coordination are key determinants of GDP (and GTP) binding affinity.

The formation of a complex between a G protein-coupled receptor (GPCR) and a heterotrimeric G protein is responsible for the majority of transmembrane signalling in response to hormones and neurotransmitters. Heterotrimeric G proteins are composed of a nucleotide-binding α -subunit (G α) and an obligate dimer of the β and γ -subunits (G $\beta\gamma$). In their inactive form, G α -subunits are bound to GDP and tightly associated with G $\beta\gamma$. G α proteins are evolutionarily related to the Ras family of G proteins but contain a small globular domain, referred to as the α -helical domain (G α sAH) (Fig. 1a). The Ras-like GTPase domain (G α sRas) contains most of the catalytic residues necessary for GTP hydrolysis, as well as the G $\beta\gamma$ and effector binding regions⁴. The structures within these regions differ between GTP- and GDP-bound states, and are appropriately termed switch regions^{5,6}.

The process of G protein activation involves formation of a complex (often referred to as a ternary complex) consisting of an agonist-occupied GPCR and a nucleotide-free G protein heterotrimer. In a companion manuscript², we present the β_2 AR–Gs crystal structure (Fig. 1b), which provides the first high-resolution snapshot of such ternary complex. However, a static view of this structure provides only limited insight into the mechanism of GDP release, the first step in G protein activation. In an effort to provide dynamic insights into the mechanism of receptor-mediated activation of Gs, we studied the

effect of agonist-bound β_2 AR on the structure of Gs using peptide amide hydrogen–deuterium exchange mass spectrometry (DXMS, see Online Methods). The rates of exchange of the amide hydrogens

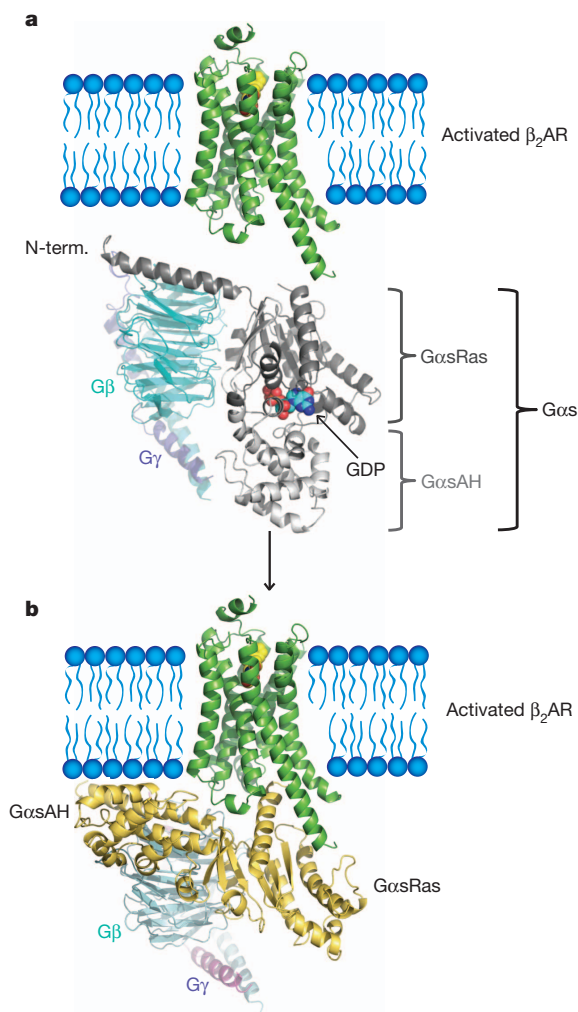


Figure 1 | Receptor-mediated activation of G proteins through promoting GDP release. Agonist (BI-167107)-bound receptor-mediated interaction with heterotrimeric G proteins promotes nucleotide (GDP) release in the G protein α -subunit. **a**, Structures of β_2 AR (green) and G protein heterotrimer (G $\alpha\beta\gamma$). The nucleotide-binding G α -subunit (grey) is composed of a Ras-homology domain (G α sRas) and an α -helical domain (G α sAH). **b**, The crystal structure of BI-167107 (agonist)-bound β_2 AR with Gs reveals large domain movements in the α -subunit (yellow) of Gs that are associated with nucleotide release.

¹Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford, California 94305, USA. ²Department of Neuroscience and Pharmacology, The Panum Institute, University of Copenhagen, 2200 Copenhagen N, Denmark. ³Department of Medicine, Biomedical Sciences Graduate Program and UCSD DXMS Proteomics Resource, University of California San Diego, La Jolla, California 92023, USA. ⁴Department of Pharmacology, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA. ⁵Department of Chemistry, University of Wisconsin, Madison, Wisconsin 53706, USA. [†]Present address: Department of Bionano Engineering, Hanyang University, Ansan 426-791, Korea.

*These authors contributed equally to this work.

in a protein are a function of the protein's thermodynamic stability, most particularly the stability of the hydrogen bonds that each amide forms in the protein's native structure. In general, an amide hydrogen involved in an intra-molecular hydrogen bond, such as those that occur in a β -strand or α -helix in a relatively stable region of the protein, exchanges more slowly than amide hydrogens forming less stable bonds^{7–12}. Changes in exchange rates reflect the propensity to formation or disruption of these bonds, thereby providing information about protein structural changes and dynamics.

We examine the amide hydrogen exchange behaviour of (1) the Gs heterotrimer alone, (2) Gs in complex with β_2 AR, (3) following uncoupling (dissociation) of the β_2 AR–Gs complex with transition state analogue, GDP/aluminium fluoride (AlF_3 or $[\text{AlF}_4]^-$), and (4) when the β_2 AR–Gs complex is exposed to GDP (Fig. 2, Supplementary Figs 2–9a). When comparing the Gs heterotrimer with the β_2 AR–Gs complex, we observe marked changes in exchange rates for G α s (Gs protein α -subunit), but little change in the exchange rates of G β or G γ

(Supplementary Figs 2–5). In addition, we do not report exchange behaviour on the β_2 AR itself because comparatively limited peptide probe sequence coverage was obtained for the receptor when we used DXMS analysis conditions optimal for the G proteins¹³.

Transitioning from the GDP-bound heterotrimer to the nucleotide-free β_2 AR–Gs complex is associated with changes in exchange rates throughout the G α subunit (Fig. 2a). There is a general increase in exchange in the segments that form the nucleotide binding pocket, consistent with a loss of polar interactions with GDP or as a result of alterations in protein structure following GDP dissociation. We observe reduced exchange at the carboxy-terminal portion of the $\alpha 5$ -helix when bound to β_2 AR (Figs 2a and 3a, b). This is not surprising, considering the crystal structures of opsin and metarhodopsin bound to a C-terminal peptide of transducin^{14–16}, and of the β_2 AR bound to Gs (ref. 2); in these three structures, the C terminus of G α subunits penetrates into the cytoplasmic core of the transmembrane bundle of GPCRs. This finding is also consistent with a plethora of

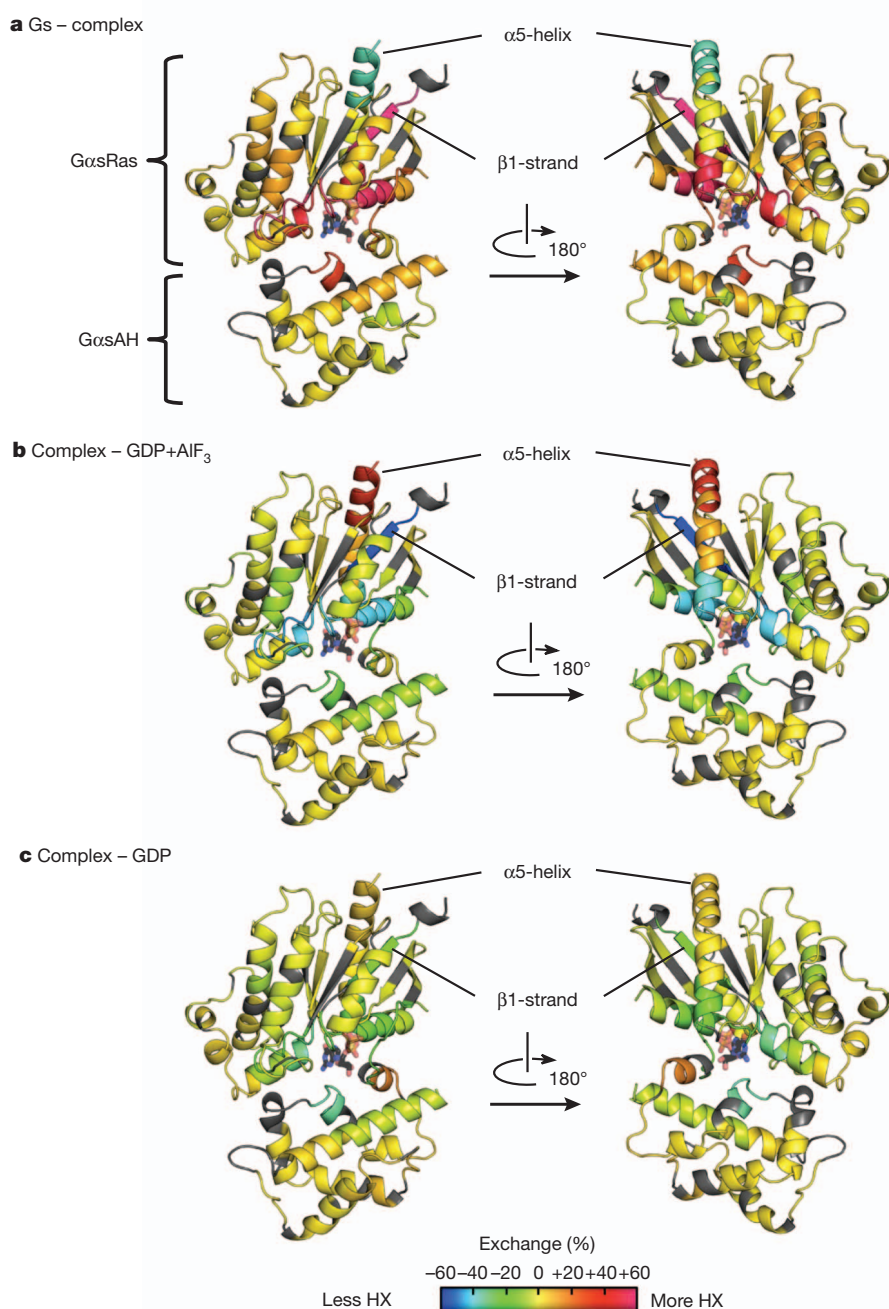


Figure 2 | DXMS reveals conformational changes in G α s when in complex with agonist-bound β_2 AR in solution. Pairwise comparisons of DXMS of G α s under different conditions.

a, Changes in DXMS of G α s following formation of the β_2 AR–Gs complex relative to G α s in the Gs heterotrimer. **b**, Changes in DXMS of G α s in the β_2 AR–Gs complex following dissociation of the complex with GDP/AlF₃. **c**, Changes in DXMS of G α s in the β_2 AR–Gs complex following the addition of GDP alone. The changes in amide hydrogen–deuterium exchange (HX), given on the colour scale (bottom) as changes in the percentage of the theoretical maximum number of deuterons incorporated per peptide, were mapped on to the crystal structure of G α s based on the GTP γ S-bound form (PDB 1AZT)²¹. Residues displaying increases (magenta) and decreases (blue) in deuterium incorporation when comparing different conditions were plotted according to the indicated heat map. Regions where no mass spectrometry data were obtained are indicated in grey. Among three time points analysed (see Supplementary Fig. 2), 100 s time points are presented.

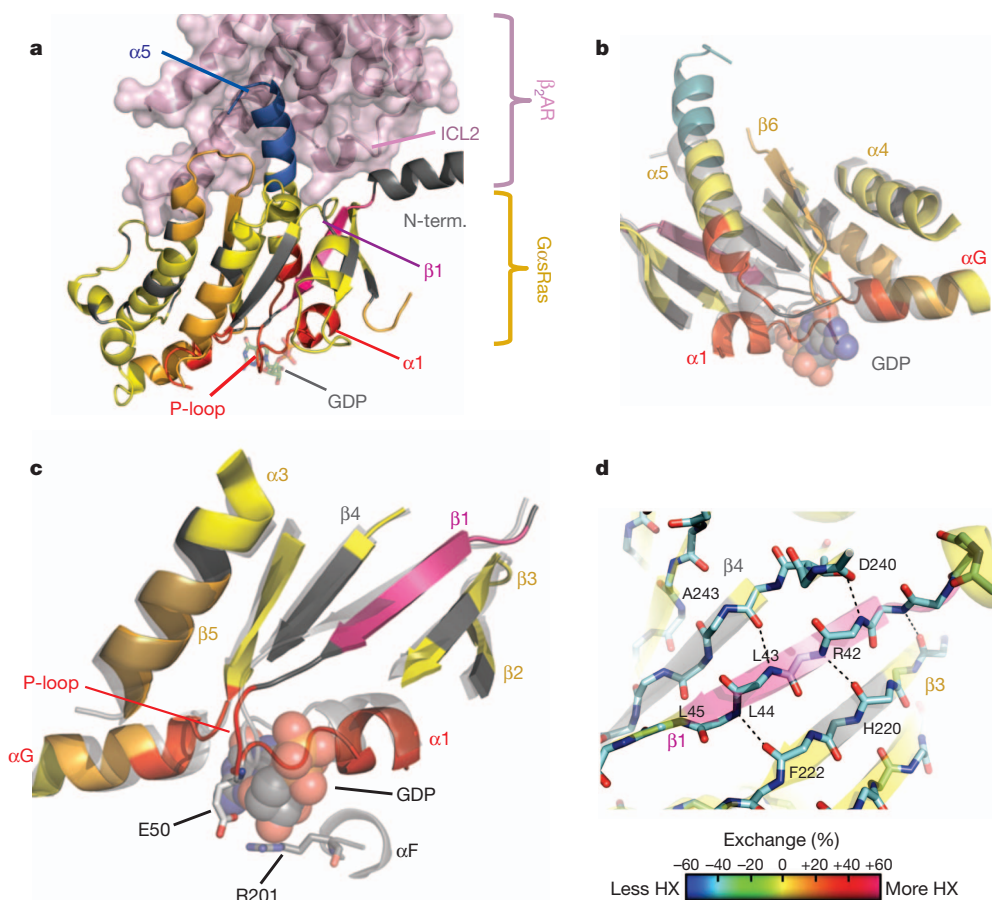


Figure 3 | Agonist-bound β_2 AR-mediated activation of GDP release revealed by DXMS. **a**, Conformational changes in the GsRas of Gs when in complex with agonist-bound β_2 AR (mauve). Highlighted are the changes in deuterium exchange (HX), according to the indicated heat map, of key regions of GsRas mapped onto the crystal structure of the complex (ref. 2). Colour scale applies to all panels, **a–d**. **b**, **c**, Comparison of the structure of Gs in complex with β_2 AR with that of Gxi (the inhibitory G protein for adenylyl cyclase bound to GDP), highlighting regions of increases or decreases in HX in the $\beta 6$ -strand- $\alpha 5$ helix (**b**) or in the $\beta 1$ -strand, P-loop and $\alpha 1$ -helix (**c**). As

above, Gxs (in **b** and **c**) is coloured according to the indicated heat map representing changes in HX when comparing Gs heterotrimer with the nucleotide-free β_2 AR-Gs complex. Residues where no mass information was obtained are coloured charcoal grey. The superimposed structure of Gxi bound to GDP (transparent grey) is based on the heterotrimeric GDP-bound form²², but illustrating the conserved glutamate (E50 in Gxs) in the GxsRas and arginine (R102 in Gxs) located in the GxsAH. **d**, Five amide hydrogens, contributed by the highly conserved 'RLLL' motif, are involved in stabilizing the $\beta 1$ -strand in a peptide fragment of Gs that displays high HX.

biochemical and biophysical evidence (reviewed in ref. 17), as well as mutagenesis studies and peptide analysis that suggest that the C terminus determines receptor coupling specificity¹⁸.

In contrast to the C-terminal region of $\alpha 5$, we observe a dramatic increase in exchange rates of peptide amides in the N-terminal segment of the $\alpha 5$ -helix, suggesting an increase in the dynamics or disordering of this segment that is not evident in the structure of the β_2 AR-Gs complex (Figs 2a, 3a, b, and Supplementary Fig. 11a). The $\beta 6$ - $\alpha 5$ loop constitutes part of the nucleotide-binding pocket. Thus, increases in exchange at the N-terminal portion of the $\alpha 5$ -helix probably reflect the sensitivity to nucleotide release (Figs 2a, 3b). The reintroduction of GDP reverses the effects observed on formation of the complex in this region (Fig. 2c). These observations are consistent with EPR spectroscopy studies suggesting movement of the $\alpha 5$ -helix of the G protein transducin on interacting with photo-activated rhodopsin¹⁹.

As expected, uncoupling Gs from β_2 AR with GDP/AlF₃ reverses all exchange changes induced by receptor binding (Fig. 2b). GDP/AlF₃ is a transition state analogue that structurally and functionally mimics GTP when bound to G proteins. Interestingly, the overall effect of adding GDP on exchange rates was similar to the effect of adding GDP/AlF₃; however, we did not observe a comparable increase in exchange at the C-terminal end of the $\alpha 5$ -helix (Fig. 2c). These results suggest that the C-terminal peptide of GDP-bound Gxs may remain

weakly coupled to the receptor, and are consistent with crosslinking experiments performed under similar conditions showing persistent interactions between β_2 AR and Gs in the presence of GDP (Supplementary Fig. 10). It is tempting to speculate that this GDP-bound β_2 AR-Gs complex may represent a 'pre-coupled' receptor complex. Nevertheless, it is likely that the GDP-bound β_2 AR-Gxs complex is not very stable, because increased exchange at the $\alpha 5$ -helix C terminus could be observed following longer exchange durations or incubations at higher (but still physiological) temperatures (see Supplementary Figs 2 and 10).

The most striking effect of β_2 AR-Gs complex formation on amide hydrogen exchange was observed in the $\beta 1$ -strand, which links the second intracellular loop (ICL2) of the agonist-bound receptor to the P-loop that coordinates the β -phosphate of GDP in Gxs (Figs 2a, 3a, c). The large increase in exchange rate in the $\beta 1$ -strand is unexpected on the basis of relatively small structural changes in $\beta 1$ - $\alpha 1$ loop observed when comparing the structures of Gxi bound to GDP with the β_2 AR-Gs complex (Fig. 3c). The relatively minor differences observed in the crystal structures may not reflect the change required for GDP release, because the crystallization process often favours the lowest energy and most stable conformation. In contrast, DXMS experimentation interrogates the entire ensemble of native-state conformations under physiological conditions (pH, salt, protein concentration and temperature)¹². Such

analysis reveals that the β 1-P-loop- α 1 region undergoes structural changes that markedly alter the stability of amide backbone hydrogen bonds.

The crystal structure of $G\alpha s$ -GTP γ S suggests that five amide hydrogens stabilize the backbone of the β 1-strand of this peptidic fragment (Fig. 3d and Supplementary Fig. 11b). Four of the five amides are contributed by the highly conserved $R_{42}LL_{45}$ motif, found in all G protein α -subunits (see alignment in Supplementary Fig. 12). DXMS analysis indicates that 4–5 hydrogens rapidly exchange in this fragment. We were unable to detect fragments of the neighbouring β 3 and β 4 strands by mass spectrometry, but speculate that alterations in exchange might also be observed in these regions. The dramatic increase in exchange in this highly conserved motif suggests that formation of the β_2 AR-Gs complex is associated with large changes in the structure and/or stability of the β -sheet formed by the β 1– β 5 strands.

The increased exchange in the β 1-strand, the P-loop and the proximal α 1-helix is compatible with the well-established mechanism of nucleotide exchange in the Ras family of G proteins. Here the β -phosphate of GDP is wedged between the highly conserved 'P-loop' and α 1-helix, and a single Mg^{2+} ion that is coordinated by residues within the P-loop. GDP release in Ras-like G proteins is manifested simply by altering Mg^{2+} coordination to the β -phosphate⁵. Although heterotrimeric $G\alpha$ proteins lack the dependence on Mg^{2+} for GDP binding, like all G proteins they display a 10^5 – 10^6 -fold preference for GDP over GMP⁶. The structural similarities between the Ras and heterotrimeric family of G proteins suggest that the mechanism for nucleotide exchange should be similar. Thus, a critical action of the

agonist-bound receptor appears to be the engagement of the N terminus of $G\alpha$, thereby altering the position and/or stability of the β 1-strand with associated changes in the structure of the P-loop and α 1-helix.

DXMS analysis also reveals regions with higher exchange that extend beyond the nucleotide-binding site and encompass the entire interface between the $G\alpha s$ AH and $G\alpha s$ Ras (Figs 2, 4). These data are consistent with crystallographic and electron microscopy evidence on the β_2 AR-Gs complex (refs 2, 3). These studies and recent double electron-electron resonance analyses²⁰ suggest that the $G\alpha s$ Ras and $G\alpha s$ AH undergo a 'clam-shell like' opening in the absence of nucleotide (see Fig. 4). Nucleotide release appears to result in re-positioning of the $G\alpha s$ AH, but the observed slow rate of exchange within the core of the domain suggests that the domain itself is intact and folded. In the GDP- and GTP γ S-bound forms of G proteins, a highly conserved glutamate residue in the P-loop (E50 in $G\alpha s$) tightly binds the highly conserved catalytic arginine residue (R201 in $G\alpha s$), which is located in the $G\alpha s$ AH (Fig. 3c and see alignment in Supplementary Fig. 12). However, the side chain of E50 in $G\alpha s$ in the receptor complex is poorly ordered in the crystal structure, displaying high *B* (temperature) factors. We propose that alterations in the P-loop structure resulting from the loss of β -phosphate coordination disrupts the E50–R201 interaction and facilitates the dissociation of $G\alpha s$ AH from $G\alpha s$ Ras.

In summary, DXMS analysis of the β_2 AR-Gs complex complements the crystal structure and provides essential additional mechanistic insights into the structural events linking GPCR-G protein complex formation and GDP release. The N and C termini of $G\alpha s$ represent the principal structural conduits between the receptor and the nucleotide binding pocket (Fig. 5). Although it is not at present possible to determine the exact sequence of events leading to nucleotide dissociation and the formation of a stable complex, we speculate that the initial interaction may involve the C terminus of GDP-bound $G\alpha s$ and agonist-bound β_2 AR. Subsequent interactions between ICL2 of the β_2 AR and the α N-helix of $G\alpha s$ are associated with GDP release through dynamic changes in the β 1-strand that involve disruption of a

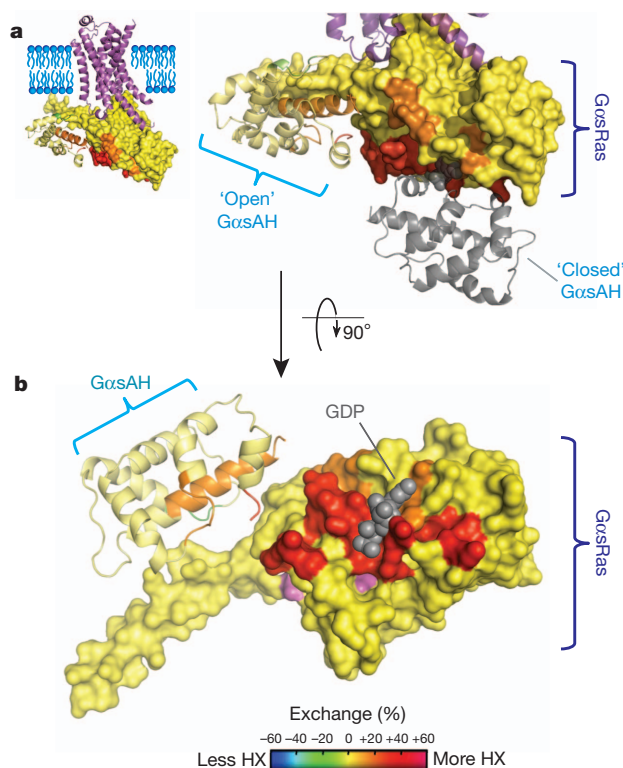


Figure 4 | Deuterium exchange at the interface between the α -helical and Ras homology domains. **a**, Changes in HX at the interface between the $G\alpha s$ AH (ribbon diagram) and $G\alpha s$ Ras (surface rendering) are mapped onto the 'open' conformation of $G\alpha s$ observed in the β_2 AR-Gs complex (inset). In this 'open' conformation, $G\alpha s$ AH and $G\alpha s$ Ras are coloured according to the indicated heat map. Also shown is a ribbon diagram of $G\alpha s$ AH in a 'closed' position (grey) similar to that observed in the crystal structure of the GDP-bound $G\alpha i$ heterotrimer. The location of the GDP binding site is shown as spheres. **b**, Surface rendering of **a** rotated back 90° to show the cytoplasmic side of the 'open' conformation of nucleotide-free $G\alpha s$.

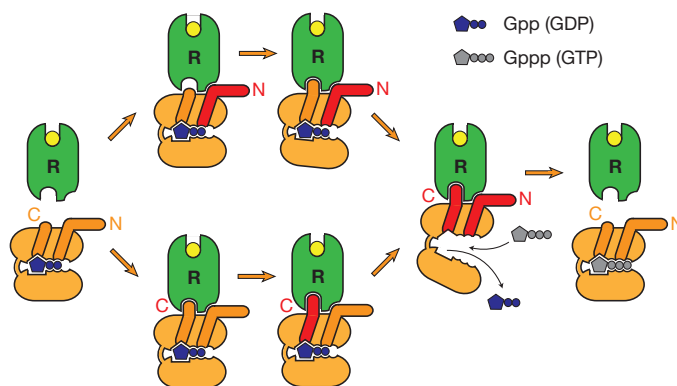


Figure 5 | Mechanism for receptor-catalysed nucleotide exchange in heterotrimeric G protein α -subunits. Shown is the step-wise dissociation of GDP from $G\alpha s$ (orange) by agonist-activated β_2 AR that involves the engagement of both the N and the C terminus of $G\alpha s$. The activation of Gs through an activated β_2 AR (green) results in GDP release and subsequent GTP binding. The activated receptor engages the C terminus of the α 5-helix of $G\alpha s$ which undergoes a rigid-body translation upward into the receptor core and reorganizes the β 6- α 5 loop, a region that participates in purine ring binding. In a simultaneous or sequential event, ICL2 of the β_2 AR engages the N terminus of the $G\alpha s$, leading to reorganization of its β 1-strand/P-loop, the loss of coordination of the β -phosphate of GDP (blue), and subsequently GDP release. The position of the N-terminal helix is aided by the $G\beta\gamma$ -subunits (not shown). The concomitant disruption of the interaction between the P-loop and $G\alpha s$ AH, primarily through the highly conserved R201 in the $G\alpha s$ AH and E50 in the P-loop, opens $G\alpha s$ AH allowing GDP to freely dissociate. Formation of the nucleotide-free form allows GTP (grey) to bind, resulting in reformation of the 'closed' conformation, and activation of the G protein through functionally dissociating from $G\beta\gamma$ and uncoupling from β_2 AR.

conserved hydrogen bond network and are coupled to structural changes in the highly conserved 'P-loop' that wraps around the β -phosphate of GDP. Thus, in a coordinated event, the agonist-bound receptor engages both the N terminus and C terminus to promote nucleotide release.

METHODS SUMMARY

DXMS studies were performed on purified Gs and β_2 AR-Gs complex in the absence or presence of guanine nucleotides. The β_2 AR-Gs complex was formed from β_2 AR and Gs protein expressed in insect (*Spodoptera frugiperda* and *Trichoplusia ni*, respectively) cells. Samples were incubated on ice for 100, 1,000 and 10,000 s in the presence of D₂O, then quenched and frozen for subsequent analysis by fragmentation and mass spectrometry. For more experimental details see Online Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 22 June; accepted 17 August 2011.

- Lefkowitz, R. J. Seven transmembrane receptors: something old, something new. *Acta Physiol.* **190**, 9–19 (2007).
- Rasmussen, S. G. *et al.* Crystal structure of the β_2 adrenergic receptor–Gs protein complex. *Nature* doi:10.1038/nature10361 (this issue).
- Westfield, G. H. *et al.* Structural flexibility of the G α s α -helical domain in the β_2 -adrenoceptor Gs complex. *Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.1113645108 (in the press).
- Sprang, S. R. G protein mechanisms: insights from structural analysis. *Annu. Rev. Biochem.* **66**, 639–678 (1997).
- Sprang, S. R. G proteins, effectors and GAPs: structure and mechanism. *Curr. Opin. Struct. Biol.* **7**, 849–856 (1997).
- Bos, J. L., Rehmann, H. & Wittinghofer, A. GEFs and GAPs: critical elements in the control of small G proteins. *Cell* **129**, 865–877 (2007).
- Englander, S. W., Mayne, L. & Krishna, M. M. Protein folding and misfolding: mechanism and principles. *Q. Rev. Biophys.* **40**, 287–326 (2007).
- Konermann, L., Tong, X. & Pan, Y. Protein structure and dynamics studied by mass spectrometry: H/D exchange, hydroxyl radical labeling, and related approaches. *J. Mass Spectrom.* **43**, 1021–1036 (2008).
- Tsutsui, Y. & Winrod, P. L. Hydrogen/deuterium exchange-mass spectrometry: a powerful tool for probing protein structure, dynamics and interactions. *Curr. Med. Chem.* **14**, 2344–2358 (2007).
- Engen, J. R. Analysis of protein conformation and dynamics by hydrogen/deuterium exchange MS. *Anal. Chem.* **81**, 7870–7875 (2009).
- Abzalimov, R. R., Kaplan, D. A., Easterling, M. L. & Kaltashov, I. A. Protein conformations can be probed in top-down HDX MS experiments utilizing electron transfer dissociation of protein ions without hydrogen scrambling. *J. Am. Soc. Mass Spectrom.* **20**, 1514–1517 (2009).
- Konermann, L., Pan, J. & Liu, Y. H. Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chem. Soc. Rev.* **40**, 1224–1234 (2011).
- Zhang, X. *et al.* Dynamics of the β_2 -adrenergic G-protein coupled receptor revealed by hydrogen-deuterium exchange. *Anal. Chem.* **82**, 1100–1108 (2010).
- Scheerer, P. *et al.* Crystal structure of opsin in its G-protein-interacting conformation. *Nature* **455**, 497–502 (2008).
- Standfuss, J. *et al.* The structural basis of agonist-induced activation in constitutively active rhodopsin. *Nature* **471**, 656–660 (2011).
- Choe, H. W. *et al.* Crystal structure of metarhodopsin II. *Nature* **471**, 651–655 (2011).
- Oldham, W. M. & Hamm, H. E. Structural basis of function in heterotrimeric G proteins. *Q. Rev. Biophys.* **39**, 117–166 (2006).
- Gilchrist, A. *et al.* Antagonists of the receptor-G protein interface block Gi-coupled signal transduction. *J. Biol. Chem.* **273**, 14912–14919 (1998).
- Oldham, W. M., Van Eps, N., Preiner, A. M., Hubbell, W. L. & Hamm, H. E. Mechanism of the receptor-catalyzed activation of heterotrimeric G proteins. *Nature Struct. Mol. Biol.* **13**, 772–777 (2006).
- Van Eps, N. *et al.* Interaction of a G protein with an activated receptor opens the interdomain interface in the α subunit. *Proc. Natl Acad. Sci. USA* **108**, 9420–9424 (2011).
- Sunahara, R. K., Tesmer, J. J., Gilman, A. G. & Sprang, S. R. Crystal structure of the adenylyl cyclase activator G α s. *Science* **278**, 1943–1947 (1997).
- Wall, M. A. *et al.* The structure of the G protein heterotrimer G $_{i\alpha 1}\beta_1\gamma_2$. *Cell* **83**, 1047–1058 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J.J.G. Tesmer and G. Skiniotis for discussions. This work was supported by an American Lung Association senior research training fellowship (RT-166882-N, to K.Y.C.), the Lundbeck Foundation (a junior group leader fellowship, to S.G.F.R.), a National Institute of General Medical Sciences (NIGMS) molecular biophysics training grant (GM008270, to B.T.D.), a National Institute of Neural Disorders and Stroke grant (NS28471, to B.K.K.), a NHLBI grant (HL071078, to B.K.K.), the Mather Charitable Foundation (B.K.K.), NIGMS grants GM083118 (to B.K.K. and R.K.S.) and GM068603 (to R.K.S.), a Michigan Diabetes Research and Training Center grant, the National Institute of Diabetes and Digestive and Kidney Diseases (P60DK-20572, to R.K.S.), the University of Michigan Biological Sciences Scholars Program (R.K.S.), and by the NIH (grants AI076961, AI081982, AI2008031, CA118595, GM20501, GM066170, GM093325 and RR029388 (to V.L.W.)).

Author Contributions K.Y.C. prepared nucleotide-treated samples for DXMS, performed hydrogen–deuterium exchange of prepared samples, analysed mass spectrometry data, and performed cross-linking studies. S.G.F.R. performed the final stages of β_2 AR purification, assisted with β_2 AR and Gs protein virus production and expression in insect cell cultures, and worked out conditions to form, stabilize and purify the β_2 AR–Gs complex using agonist BI-167107 and detergent MNG-3. T.L. was responsible for the overall data analysis. S.L. ran prepared samples on the mass spectrometer. B.T.D. managed Gs heterotrimer subunit virus production and titration, expressed and purified Gs protein, and, with K.Y.C. and T.L., analysed mass spectrometry data. P.S.C. provided MNG-3 detergent for stabilization of the β_2 AR–Gs complex. D.C. assisted with Gs heterotrimer expression and purification. B.K.K. was responsible for overall project strategy and management, and assisted with manuscript preparation. V.L.W. oversaw design, execution and data analysis of DXMS experiments, and assisted with manuscript preparation. R.K.S. supervised Gs protein production, provided ideas and insights into Gs structure and function, and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.K.S. (sunahara@umich.edu), B.K.K. (kobilka@stanford.edu) or V.L.W. (vwoods@ucsd.edu).

METHODS

Expression and purification of β_2 AR. An β_2 AR construct truncated at position 365 (β_2 AR-365) was expressed in Sf9 insect (*Spodoptera frugiperda*) cell cultures (grown in ESF 921 medium, Expression Systems) infected with recombinant baculovirus (BestBac, Expression Systems), and solubilized in n-dodecyl- β -D-maltoside (DDM) according to methods described previously²³. M1 Flag affinity chromatography (Sigma) served as the initial purification step followed by alprenolol-Sepharose chromatography for selection of functional receptor. A subsequent M1 Flag affinity chromatography step was used to exchange receptor-bound alprenolol for high-affinity agonist BI-167107. The agonist-bound receptor was eluted, dialysed against buffer (20 mM HEPES, pH 7.5, 100 mM NaCl, 0.1% DDM and 10 μ M BI-167107), treated with PNGaseF and lambda phosphatase (New England Biolabs), and concentrated to approximately 50 mg ml⁻¹ with a 50 kDa molecular weight cut-off Millipore concentrator.

Expression and purification of Gs heterotrimer. Bovine G α_s short, His₆-rat G β_1 , and bovine G γ_2 were co-expressed in HighFive insect (*Trichoplusia ni*) cells (Invitrogen) grown in Insect Xpress serum-free media (Lonza). Cultures were infected with three separate *Autographa californica* nuclear polyhedrosis viruses, each containing the gene for one of the G protein subunits at a 1:1 multiplicity of infection. Infected cells were harvested and resuspended in 75 ml lysis buffer (50 mM HEPES pH 8.0, 65 mM NaCl, 1.1 mM MgCl₂, 1 mM EDTA, 1 \times PTT (35 μ g ml⁻¹ phenylmethanesulphonyl fluoride, 32 μ g ml⁻¹ tosyl-phenylalanyl-chloromethyl ketone, 32 μ g ml⁻¹ tosyl-lysyl-chloromethyl ketone), 1 \times LS (3.2 μ g ml⁻¹ leupeptin and 3.2 μ g ml⁻¹ soybean trypsin inhibitor), 5 mM β -ME, and 10 μ M GDP) per litre of culture volume. The suspension was pressurized with 600 p.s.i. N₂ for 40 min in a nitrogen cavitation bomb (Parr Instrument Company). After depressurization, the lysate was centrifuged to remove nuclei and unlysed cells, and then ultracentrifuged at 180,000g for 40 min. The pelleted membranes were resuspended in 30 ml wash buffer (50 mM HEPES, pH 8.0, 50 mM NaCl, 100 μ M MgCl₂, 1 \times PTT, 1 \times LS, 5 mM β -ME, 10 μ M GDP) per litre culture volume using a Dounce homogenizer and centrifuged again at 180,000g for 40 min. The washed pellet was resuspended in a minimal volume of wash buffer and flash frozen with liquid nitrogen.

The frozen membranes were thawed and diluted to a total protein concentration of 5 mg ml⁻¹ with fresh wash buffer. Sodium cholate detergent was added to the suspension at a final concentration of 1.0%, MgCl₂ was added to a final concentration of 5 mM, and 0.05 mg of purified protein phosphatase 5 (PP5, prepared from *E. coli* culture) was added per litre of culture volume. PP5 cDNA was generously provided by M. Chinkers (University of Southern Alabama). The sample was stirred on ice for 40 min, and then centrifuged at 180,000g for 40 min to remove insoluble debris. The supernatant was diluted fivefold with Ni-NTA load buffer (20 mM HEPES, pH 8.0, 363 mM NaCl, 1.25 mM MgCl₂, 6.25 mM imidazole, 0.2% Anzergent 3-12, 1 \times PTT, 1 \times LS, 5 mM β -ME, 10 μ M GDP), taking care to add the buffer slowly to avoid reducing the cholate concentration below its critical micelle concentration too quickly. 3 ml of Ni-NTA resin (Qiagen) pre-equilibrated in Ni-NTA wash buffer 1 (20 mM HEPES pH 8.0, 300 mM NaCl, 2 mM MgCl₂, 5 mM imidazole, 0.2% sodium cholate, 0.15% Anzergent 3-12, 1 \times PTT, 1 \times LS, 5 mM β -ME, 10 μ M GDP) per litre culture volume was added and the sample was stirred on ice for 20 min. The resin was collected into a gravity column and washed with 4 \times column volumes of Ni-NTA wash buffer 1, Ni-NTA wash buffer 2 (20 mM HEPES, pH 8.0, 50 mM NaCl, 1 mM MgCl₂, 10 mM imidazole, 0.15% Anzergent 3-12, 0.1% DDM, 1 \times PTT, 1 \times LS, 5 mM β -ME, 10 μ M GDP), and Ni-NTA wash buffer 3 (20 mM HEPES, pH 8.0, 50 mM NaCl, 1 mM MgCl₂, 5 mM imidazole, 0.1% DDM, 1 \times PTT, 1 \times LS, 5 mM β -ME, 10 μ M GDP). The protein was eluted with Ni-NTA elution buffer (20 mM HEPES, pH 8.0, 40 mM NaCl, 1 mM MgCl₂, 200 mM imidazole, 0.1% DDM, 1 \times PTT, 1 \times LS, 5 mM β -ME, 10 μ M GDP). Protein-containing fractions were pooled and MnCl₂ was added to a final concentration of 100 μ M. Purified lambda protein phosphatase (prepared from *E. coli* culture) was added to the elute (50 μ g per litre of culture volume, final) and incubated on ice with stirring for 30 min. The elute was passed through a 0.22 μ m filter and loaded directly onto a MonoQ HR 16/10 column (GE Healthcare) equilibrated in MonoQ buffer A (20 mM HEPES, pH 8.0, 50 mM NaCl, 100 μ M MgCl₂, 0.1% DDM, 5 mM β -ME, 1 \times PTT). The column was washed with 150 ml buffer A at 5 ml min⁻¹ and bound proteins were eluted over 350 ml with a linear gradient up to 28% MonoQ buffer B (same as buffer A except with 1 M NaCl). Fractions were collected in tubes spotted with enough GDP to make a final concentration of 10 μ M. The Gs containing fractions were concentrated to 2 ml using a stirred ultrafiltration cell with a 10 kDa NMWL regenerated cellulose membrane (Millipore). The concentrated sample was run on a Superdex 200 prep grade XK 16/70 column (GE Healthcare) equilibrated in S200 buffer (20 mM HEPES, pH 8.0, 100 mM NaCl, 1.1 mM MgCl₂, 1 mM EDTA, 0.012% DDM, 100 μ M TCEP, 2 μ M GDP). The fractions containing pure Gs were pooled, glycerol was added to 10% final concentration, and then the protein was

concentrated to at least 10 mg ml⁻¹ using a 30 kDa MWCO regenerated cellulose Amicon centrifugal ultrafiltration device. The concentrated sample was then aliquoted, flash frozen, and stored at -80 $^{\circ}$. A typical yield of final, purified Gs heterotrimer from 8 l of cell culture volume was 6–8 mg.

Complex formation and purification. For reasons of efficiency and higher protein quality, the receptor-G protein (β_2 AR-365-Gs heterotrimer) complex that was used for DXMS and EM (ref. 3) experimentation was derived from preparations also intended for protein crystallography². Heterotrimeric Gs protein was mixed at approximately 100 μ M concentration with BI-167107 bound β_2 AR-365 in molar excess (approximately 130 μ M) in 2 ml buffer (10 mM HEPES pH 7.5, 100 mM NaCl, 0.1% DDM, 1 mM EDTA, 3 mM MgCl₂, 10 μ M BI-167107) and incubating for 3 h at room temperature (25 $^{\circ}$ C). To obtain a nucleotide free β_2 AR-Gs complex, apyrase (25 mU ml⁻¹, NEB) was added after 90 min to hydrolyse residual GDP released from G α_s on binding to the receptor. The complex was exchanged into MNG-3 by adding the β_2 AR-Gs complex mixture (2 ml) to 8 ml buffer (20 mM HEPES pH 7.5, 100 mM NaCl, 10 μ M BI-167107) containing 1% MNG-3 for 1 h at room temperature. The β_2 AR-Gs complex was loaded onto a M1 Flag resin and washed in buffer (20 mM HEPES, pH 7.5, 100 mM NaCl, 10 μ M BI-167107 and 3 mM CaCl₂) containing 0.2% MNG-3 and eluted in buffer (20 mM HEPES, pH 7.5, 100 mM NaCl, 10 μ M BI-167107, 5 mM EDTA and 200 μ g ml⁻¹ Flag peptide) containing 0.02% MNG-3. The β_2 AR-Gs complex was purified following a final size exclusion chromatography step using a Superdex 200 10/300 GL column (GE Healthcare) equilibrated with buffer containing 0.02% MNG-3, 10 mM HEPES pH 7.5, 100 mM NaCl, 10 μ M BI-167107 and 100 μ M TCEP. Peak fractions were pooled and concentrated to approximately 90 mg ml⁻¹ with a 100 kDa MWCO Viva-spin concentrator. A representative elution profile of the complex on a Superdex 200 column can be found in Supplementary Fig. 1.

Nucleotide treatment. β_2 AR-Gs complex prepared as described above was diluted to 10 mg ml⁻¹ in buffer containing 0.02% MNG-3, 20 mM HEPES, pH 7.5, 100 mM NaCl, 10 μ M BI-167107 and 100 μ M TCEP. The β_2 AR-Gs complex was incubated with either 1 mM GDP or 1 mM GDP/AlF₃ for 2 h at room temperature.

Deuterium exchange mass spectrometry. 1.5 μ l of β_2 AR-Gs complex (10 mg ml⁻¹) or 1.5 μ l of Gs heterotrimer (7 mg ml⁻¹) was mixed with 4.5 μ l of D₂O buffer (20 mM HEPES, pH 7.5, 100 mM NaCl, 10 μ M BI-167107, 100 μ M TCEP, 0.0015% MNG-3 in D₂O) and incubated for 100, 1,000 and 10,000 s on ice. At the indicated times, the sample was quenched by 15 μ l of ice-cold quench solution (0.1 M NaH₂PO₄, 20 mM TCEP, 16.6% glycerol, pH 2.4), immediately frozen on dry ice, and stored at -80 $^{\circ}$ C. Non-deuterated control was prepared in H₂O buffer (20 mM HEPES, pH 7.5, 100 mM NaCl, 10 μ M BI-167107, 100 mM TCEP, 0.0015% MNG-3 in H₂O), mixed with quench solution, and snap-frozen on dry ice. Samples were thawed and immediately passed through an immobilized porcine pepsin column (16 μ l bed volume) at a flow rate of 20 μ l min⁻¹ of 0.05% trifluoroacetic acid. Peptide fragments were collected contemporaneously on a C18 trap column for desalting and separated by a Magic C18AQ column (Michrom BioResources Inc.) using a linear gradient of acetonitrile from 6.4% to 38.4% over 30 min. Mass spectrometric analysis was performed using LCQ Classic mass spectrometer (Thermo Finnigan), with capillary temperature of 200 $^{\circ}$ C. Deuterium quantification data were collected in MS1 profile mode, and peptide identification data were collected in data dependent MS/MS mode. Recovered peptide identification and analysis were carried out using DXMS Explorer (Sierra Analytics Inc.), specialized software for processing DXMS data^{24–26}.

Before the performance of studies with deuterated proteins, extensive preliminary 'fragmentation tuning' studies were performed with non-deuterated protein mixtures. Proteins were acidified, supplemented with various concentrations of denaturants (guanidine-HCl, 0.0–4.0 M), supplemental detergents (DDM, 0.0 and 0.1%) and reductants (TCEP, 0.015–0.5 M); incubated in such solutions for varying times at 0 $^{\circ}$ C; and then subjected to on-line solid-state pepsin proteolysis (45 s contact time at either 0 $^{\circ}$ C or room temperature) and then analysed by LC-MS at 0 $^{\circ}$ C as described above. The conditions that produced the most comprehensive fragmentation map for the G protein component of the mixture (described and employed above) were found to be markedly different from those that gave the best fragmentation map for the receptor protein component. Addition of supplemental DDM (0.1%), long incubations in concentrations of denaturants (4.0 M), or digestion at room temperature were required to obtain even a partial fragmentation map of the receptor. Compared to the G protein, the receptor is more highly disulphide cross-linked, and hydrophobic in character, probably accounting for the requirement for a more aggressive denaturation and protease digestion, and supplementation with additional detergent to keep hydrophobic peptides in solution. These several conditions degraded the comprehensiveness and resolution of the G protein fragmentation map, and considerably increased the deuterium losses of the G protein peptides after exchange quench when compared to the G

protein-optimal conditions. Given this, our deuteration analysis was subsequently directed exclusively to the G protein component of the mixture employing analysis conditions optimal for them.

23. Kobilka, B. K. Amino and carboxyl terminal modifications to facilitate the production and purification of a G protein-coupled receptor. *Anal. Biochem.* **231**, 269–271 (1995).
24. Mendillo, M. L. *et al.* A conserved MutS homolog connector domain interface interacts with MutL homologs. *Proc. Natl Acad. Sci. USA* **106**, 22223–22228 (2009).
25. Chen, H. *et al.* Allosteric inhibition of complement function by a staphylococcal immune evasion protein. *Proc. Natl Acad. Sci. USA* **107**, 17621–17626 (2010).
26. Li, S. *et al.* Mechanism of intracellular cAMP sensor Epac2 activation: cAMP-induced conformational changes identified by amide hydrogen/deuterium exchange mass spectrometry (DXMS). *J. Biol. Chem.* **286**, 17889–17897 (2011).

The structure and catalytic mechanism of a poly(ADP-ribose) glycohydrolase

Dea Slade^{1,2*}, Mark S. Dunstan^{3*}, Eva Barkauskaite¹, Ria Weston¹, Pierre Lafite⁴, Neil Dixon³, Marijan Ahel⁵, David Leys³ & Ivan Ahel¹

Post-translational modification of proteins by poly(ADP-ribose)ylation regulates many cellular pathways that are critical for genome stability, including DNA repair, chromatin structure, mitosis and apoptosis¹. Poly(ADP-ribose) (PAR) is composed of repeating ADP-ribose units linked via a unique glycosidic ribose-ribose bond, and is synthesized from NAD by PAR polymerases^{1,2}. PAR glycohydrolase (PARG) is the only protein capable of specific hydrolysis of the ribose-ribose bonds present in PAR chains; its deficiency leads to cell death^{3,4}. Here we show that filamentous fungi and a number of bacteria possess a divergent form of PARG that has all the main characteristics of the human PARG enzyme. We present the first PARG crystal structure (derived from the bacterium *Thermomonospora curvata*), which reveals that the PARG catalytic domain is a distant member of the ubiquitous ADP-ribose-binding macrodomain family^{5,6}. High-resolution structures of *T. curvata* PARG in complexes with ADP-ribose and the PARG inhibitor ADP-HPD, complemented by biochemical studies, allow us to propose a model for PAR binding and catalysis by PARG. The insights into the PARG structure and catalytic mechanism should greatly improve our understanding of how PARG activity controls reversible protein poly(ADP-ribose)ylation and potentially of how the defects in this regulation are linked to human disease.

Whereas the mechanism and functional aspects of PAR synthesis have been extensively characterized, the present understanding of the PAR-degradation pathways catalysed by PARG is comparatively poor. Until now, the catalytic fold and the mechanism of PARG-mediated hydrolysis have remained unknown. Our homology searches revealed that the majority of fungal genomes lack close mammalian/canonical PARG homologues, but instead possess a divergent PARG-like protein, annotated as DUF2263. DUF2263 protein sequences contain the PARG signature (GGG-X₆₋₈-QEE)⁷, which includes the previously identified key residues: two consecutive glutamates⁷ (Fig. 1a, black asterisk) and a glycine⁸ (Fig. 1a, grey asterisk). In addition, a tyrosine residue (Fig. 1a, black cross) implicated in binding to the PARG inhibitor ADP-HPD⁹ is also present in DUF2263. The DUF2263 orthologue is found also in other eukaryotes lacking canonical PARGs such as the rotifer *Adineta vaga* (Fig. 1b), where it is fused to several copies of the PAR-binding zinc finger¹⁰. As both filamentous fungi and *A. vaga* have functional PARP orthologues and an active PAR metabolism^{11,12}, we postulated that DUF2263 was likely to be a functional PARG in these organisms. Surprisingly, although bacteria are historically thought to be devoid of PAR metabolism, certain bacterial species possess both PARP (closely related to PARP1; ref 13) and DUF2263 genes, whereas some only contain the DUF2263 homologue (Fig. 1a, b). To analyse the biochemical activities of the corresponding bacterial PARP and PARG proteins, we purified His₆-tagged *Herpetosiphon aurantiacus* PARP and DUF2263. *H. aurantiacus* PARP was able to synthesize PAR from NAD, as revealed by antibodies recognizing PAR (Fig. 1c)

and by the analysis of PAR ladders using sequencing gels (Supplementary Fig. 1). Analogous to the human PARP1 enzyme, *H. aurantiacus* PARP was sensitive to the PARP inhibitor KU-0058948 and required DNA for its activation (Fig. 1c). Notably, poly(ADP-ribose)ylation produced by *H. aurantiacus* PARP is efficiently hydrolysed by the *H. aurantiacus* DUF2263 protein, demonstrating that DUF2263 is a functional PARG (Fig. 1c) and suggesting the presence of functional PAR metabolism in bacteria.

We extended our analyses to other bacterial and fungal PARGs (Fig. 1a, b and Supplementary Fig. 2) using established PAR substrates formed by the activity of the well-characterized human PARP1 enzyme. Canonical human and *Entamoeba dispar* PARGs served as controls. The *in vitro* activity of the purified PARG enzymes was evaluated in five ways: (1) an assay that measures the loss of PAR from PARP1-modified histones; (2) a western-blot analysis of the hydrolysis of PAR from the auto-modified human PARP1 using anti-PAR antibodies; (3) an SDS-polyacrylamide gel electrophoresis (SDS-PAGE) analysis of the [³²P]-PAR auto-modified human PARP1; (4) an analysis of the release of ADP-ribose from PAR either by thin-layer chromatography (TLC); or (5) by liquid chromatography-mass spectrometry (LC/MS) (Fig. 2a–d and Supplementary Fig. 4). The quantification of PAR hydrolysis by the PARG assay shows that bacterial and fungal PARGs are highly active in hydrolysing PAR *in vitro*; their catalytic potency is comparable to the *E. dispar* and human PARG enzymes (Fig. 2a). Moreover, the mode of action of these proteins and mutant versions on poly(ADP-ribose)ylated PARP1 shows the same pattern observed for the canonical PARGs. First, they are all inactivated by the mutation of a key glutamate residue (Fig. 2b, d) and by treatment with the PARG inhibitor ADP-HPD (Fig. 2c). Second, they are all active on ribose-ribose bonds between the two ADP-ribose units, but they are not efficient on mono(ADP-ribose)ylated protein substrates with the ADP-ribose unit linked directly to the PARP1 (Fig. 2c) or the PARP10 protein (Supplementary Fig. 3). Last, TLC and LC/MS analyses confirmed that the main product of DUF2263-type PARGs is ADP-ribose (Fig. 2d and Supplementary Fig. 4). Collectively, these results demonstrate that bacterial and fungal PARG proteins possess genuine PARG activity *in vitro*.

To test whether bacterial and fungal PARG proteins can act as PARGs *in vivo*, we used the *Saccharomyces cerevisiae* system expressing human PARP1 (ref. 14). Yeast cells, naturally deficient in PAR metabolism, can readily express human PARP1, resulting in poly(ADP-ribose)ylation of various protein substrates¹⁴ (Fig. 2e). When divergent PARG proteins were co-expressed with human PARP1, we observed that the poly(ADP-ribose)ylation levels in yeast extracts diminished, indicating an efficient hydrolysis of PAR *in vivo* (Fig. 2e and Supplementary Fig. 5).

Although the first PARG gene was cloned in 1997 (ref. 15), until now there has been no structural data available for a PARG enzyme. To gain an insight into the PARG structure and mechanism, we solved the

¹Cancer Research UK, Paterson Institute for Cancer Research, University of Manchester, Wilmslow Road, Manchester M20 4BX, UK. ²Université de Paris-Descartes, Faculté de Médecine, INSERM U1001, 156 rue de Vaugirard, 75015 Paris, France. ³Manchester Interdisciplinary Biocentre, Princess Street 131, Manchester M1 7DN, UK. ⁴ICOA – UMR CNRS 6005 Université d'Orléans, Rue de Chartres, F-45067 Orléans, France. ⁵Rudjer Boskovic Institute, Bijenicka 54, HR-10000 Zagreb, Croatia.

*These authors contributed equally to this work.

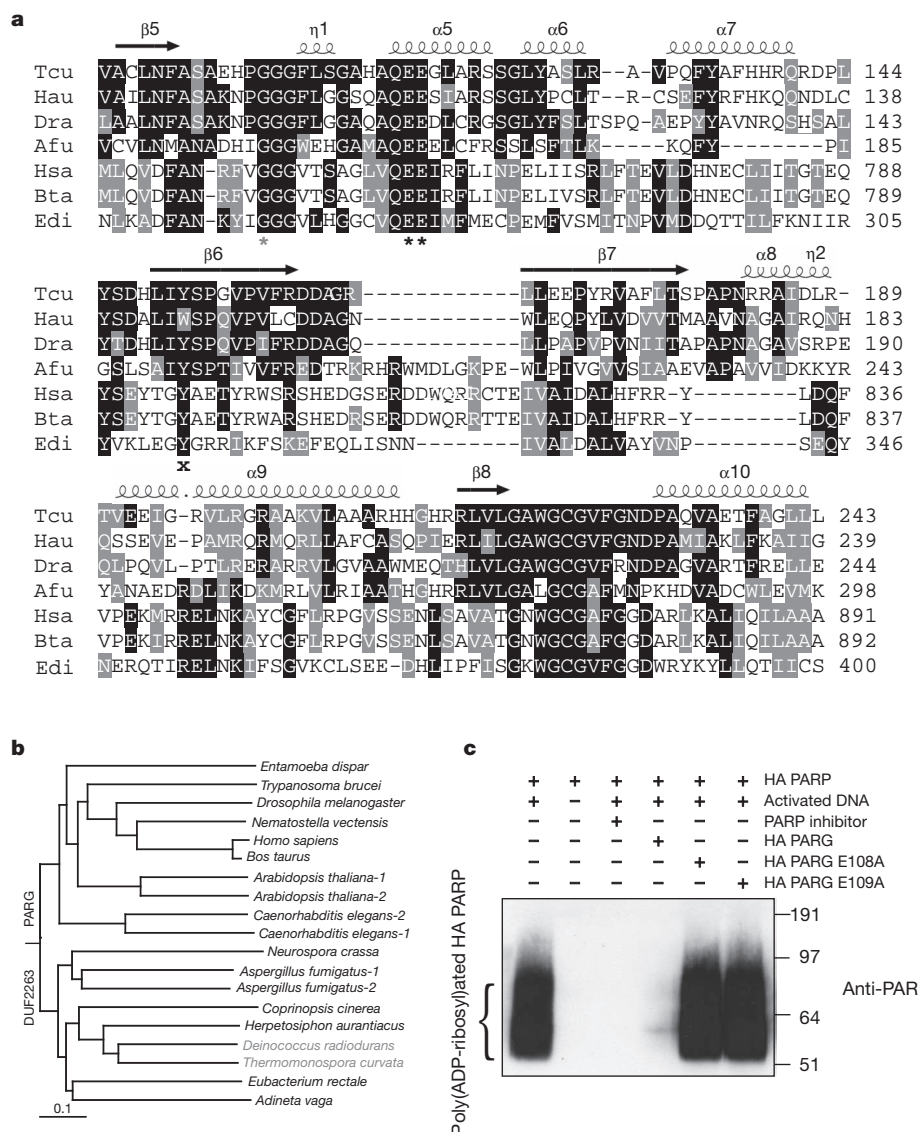


Figure 1 | Phylogeny and functional relationship between DUF2263 and canonical-type PARGs. a, Multiple sequence alignment of different DUF2263 and PARG proteins from *T. curvata* (Tcu), *H. aurantiacus* (Hau), *Deinococcus radiodurans* (Dra), *Aspergillus fumigatus* (Afu), *Homo sapiens* (Hsa), *Bos taurus* (Bta) and *E. dispar* (Edi). The two catalytic glutamates, a conserved glycine and tyrosine are marked with black asterisks, a grey asterisk and a black

T. curvata PARG ligand-free crystal structure to 1.5 Å in addition to ADP-ribose and ADP-HPD inhibitor-bound structures (Supplementary Table 1). Comparison with available structures reveals that *T. curvata* PARG consists of an ADP-ribose-binding macrodomain fold with an amino-terminal extension (residues 3–67; Fig. 3c). Similar macrodomain structures are found in the *Escherichia coli* YmdB protein (PDB accession 1SPV; Z score 16.2, root mean squared deviation (r.m.s.d.) 2.1 Å over 161 C α atoms, 19% identity), an *Archaeoglobus fulgidus* macrodomain protein (PDB accession 2BFQ; Z score 15.6, r.m.s.d. 2.8 Å over 164 C α atoms, 16% identity)⁵ and the histone variant MACROH2A1.1 (PDB accession 1ZR3; Z score 15.0, 2.3 Å r.m.s.d. over 157 C α atoms, 12% identity)¹⁶. The diphosphate-binding loop that flanks one side of the ADP-ribose-binding cavity is highly conserved between PARG and other macrodomain structures (in magenta, Fig. 3a, c). In contrast, the opposite side of the PARG ADP-ribose-binding cavity is lined by a stretch of amino acids corresponding to the PARG-specific GGG-X_{6–8}-QEE signature sequence⁷ (in yellow, Fig. 3a–c). The structural alignment reveals this PARG-specific loop is inserted into the macrodomain fold to accommodate

cross, respectively. Secondary structural elements from the Tcu PARG structure are indicated above. **b**, YmdB-rooted phylogenetic tree of PARGs implied by the neighbour-joining method. Organisms devoid of PARG are marked in grey. **c**, *H. aurantiacus* (HA) PARG and PARG enzymes are active as shown by western blotting with anti-PARG antibodies. The positions of protein markers are indicated on the right margin (in kDa).

the Glu 115 side chain that projects into the PARG active site (Fig. 3c, d). Owing to the PARG-specific loop, it seems that only PARGs but not other macrodomain proteins can hydrolyse PAR (Supplementary Fig. 6).

The ligand–PARG complex structures reveal that ADP- α -ribose occupies a position similar to that observed in other ADP-ribose–macrodomain complex structures⁵ (Fig. 3d and Supplementary Fig. 7). A minor rearrangement can be observed for Val 226–Phe 227 upon ADP-ribose binding. The Phe 227 side chain is in close contact with the ribose'' moiety of ADP-ribose, ensuring a close distance between the ribose'' O4 and one of the phosphate groups within the constraints of the active site. The ribose'' moiety also forms hydrogen bonds with Glu 114 (via the 2'-OH group) and Glu 115 (via the 1'-OH group) (Fig. 3d). With the exception of water-mediated polar contacts, few direct contacts can be observed with the adenosine moiety. In contrast to the ribose'' moiety, the adenosine ribose' moiety is markedly less accessible. The presence of any substituents on the ribose' 2'-OH group (such as another ADP-ribose) would require a major reorientation of the PARG carboxy-terminal α -helix (see Fig. 4a) with concomitant

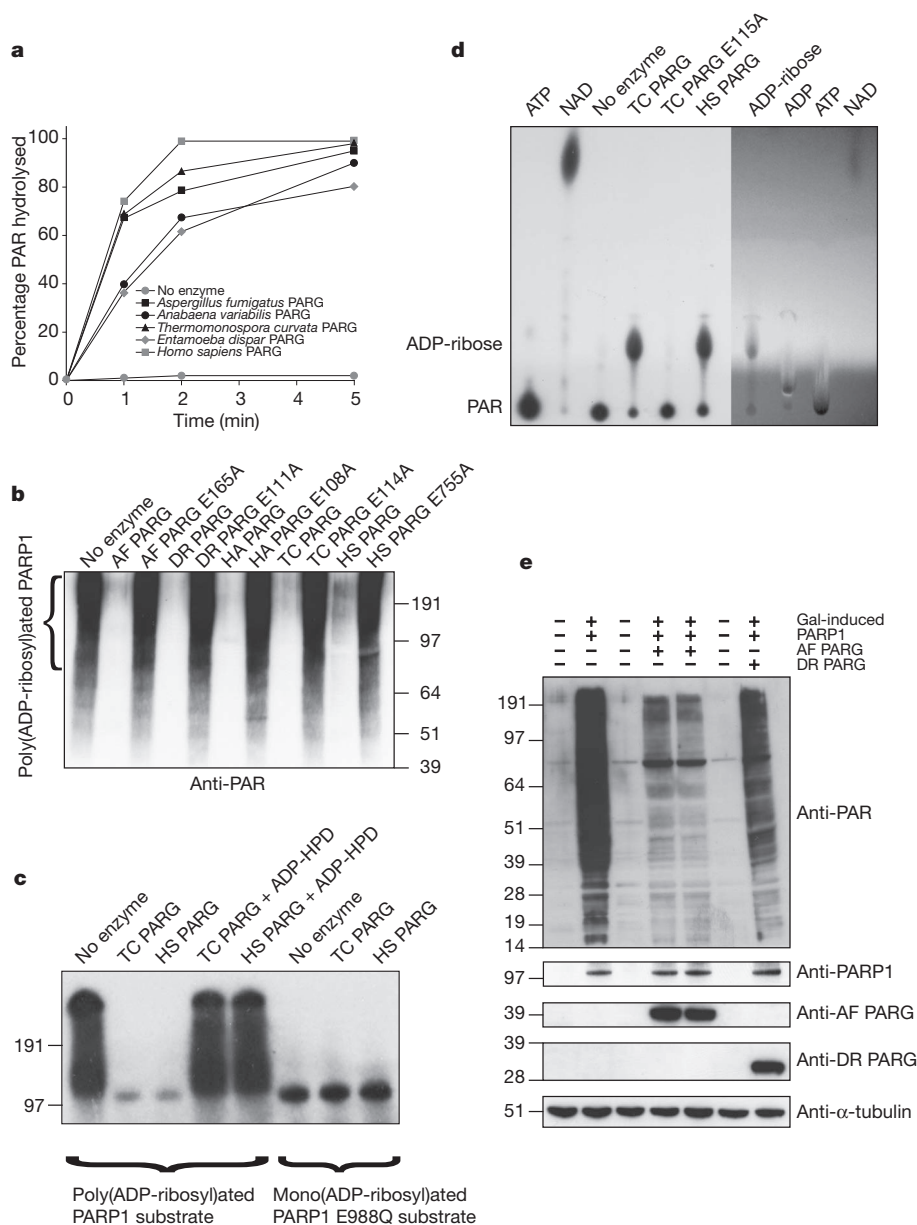


Figure 2 | PAR hydrolytic activities of divergent and canonical PARGs. **a**, A colorimetric PARG assay. **b**, Analysis of the hydrolysis of the PARP1-generated PAR substrate by anti-PAR antibodies. AF, *A. fumigatus*; AV, *A. variabilis*; DR, *D. radiodurans*; HA, *H. aurantiacus*; HS, *H. sapiens*; ED, *E. dispar*; TC, *T. curvata*. The positions of protein markers are indicated on the margins of panels **b**, **c** and **e** (in kDa). **c**, SDS-PAGE-based assay with [32 P]-automodified

PARP1 substrate. PARGs are inhibited by ADP-HPD. **d**, TLC analysis of PARG activity on the [32 P]-PAR substrate. The right side of the TLC plate was visualized by shadowing under ultraviolet light. **e**, Heterologously expressed *A. fumigatus* and *D. radiodurans* PARG hydrolyse PAR in human PARP1-expressing budding yeast cells.

exposure of the hydrophobic core of the macrodomain. This suggests that bacterial PARG binds the PAR terminus and acts as an exoglycosidase.

The steric restraints imposed by the PARG active site form a basis for the modelling of an $\alpha(1'' \rightarrow 2')$ O-glycosidic linkage with an additional ADP-ribose group to provide further insight into PAR binding (Fig. 4a, b). Our models suggest that the $(n-1)$ adenine moiety is partially enclosed by Ala 110, Ala 112 and Val 226 while the $(n-1)$ ribose' is in close contact with Ser 98 and Gly 104 (Fig. 4b). However, molecular dynamics simulations suggest that the enzyme lacks any obvious binding sites for the $(n-1)$ phosphate and ribose'' moieties, as exemplified by the considerable conformational freedom of the latter with respect to the terminal ADP-ribose and the associated $(n-1)$ ribose'. This suggests that bacterial PARG does not specifically bind additional PAR elements.

The *T. curvata* PARG complex with ADP-ribose (Fig. 3d) and the corresponding PAR-PARG model (Fig. 4b) allow us to propose a mechanism for PARG catalysis (Fig. 4c). Binding of the PAR terminus positions the key ribose-ribose O-glycosidic linkage in direct hydrogen-bonding contact with Glu 115, while constraining the conformation of the terminal ribose''. Formation of a putative oxocarbenium intermediate is supported by the protonation of the $(n-1)$ PAR ribose' 2'-OH leaving group via Glu 115, and by the stabilization of the positively charged oxocarbenium through close proximity with the terminal diphosphate group. A water molecule is ideally positioned to attack the oxocarbenium intermediate, activated through concomitant deprotonation by Glu 115. This leads to the release of ADP- β -ribose'' and $(n-1)$ PAR.

This mechanism is supported by mutagenesis and binding studies. Although mutations of both Glu 114 and Glu 115 render the enzyme

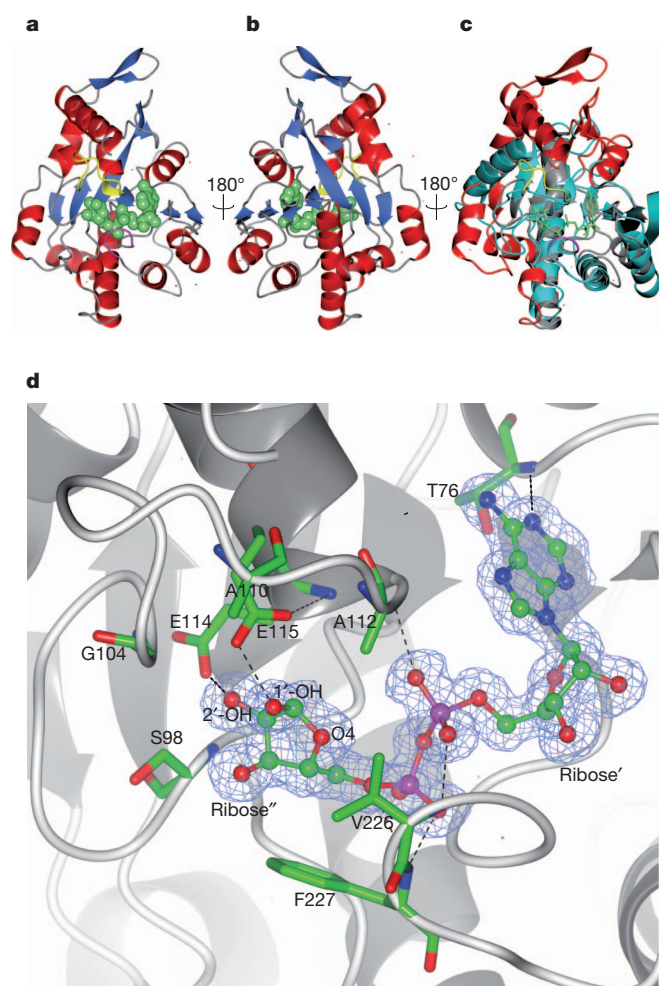


Figure 3 | *T. curvata* PARG crystal structure in complex with ADP-ribose. **a**, **b**, Overall fold of PARG with α -helices depicted in red, β -sheets in blue and bound ADP-ribose in green spheres. The PARG-specific catalytic loop is shown in yellow, and the diphosphate-binding loop in magenta. **c**, Overlay of the *T. curvata* PARG structure with a representative macrodomain structure (PDB accession 2BFQ, in cyan). The PARG-specific N-terminal extension and additional PARG structural elements are highlighted in red. Structural features that are similar to 2BFQ and other macrodomains are shown in grey. **d**, Detailed view of ADP-ribose bound in the PARG active site. ADP-ribose is shown with the corresponding $2F_o - F_c$ omit map density contoured at 1.2σ in blue. Key active site residues are represented in atom coloured sticks with hydrogen bonds indicated by black dotted lines.

inactive without disruption to the overall PARG fold (Fig. 4d and Supplementary Fig. 8), the isothermal titration calorimetry (ITC) ADP-ribose binding studies reveal that mutation of Glu 115 into an alanine does not affect binding, whereas the Glu114Ala mutation leads to an approximately tenfold decrease in binding affinity (Supplementary Fig. 9). This supports the notion that the role of Glu 114 is confined to substrate binding/orientation as opposed to the acid-base catalysis proposed for Glu 115. The mutation of Phe 227 implicated in positioning the terminal ribose'' also renders the enzyme inactive (Fig. 4d). Mutations of Ser 98 and Val 226, implicated by our model in binding the $(n - 1)$ ADP-ribose, greatly diminish *T. curvata* PARG activity (Fig. 4d). Importantly, mutations of the corresponding catalytic residues in human PARG have a similar effect on the enzymatic activity, which suggests a universal catalytic mechanism for bacterial and mammalian PARGs (Fig. 4d).

These data provide the first insight into PARG structure and catalytic mechanism, and improve our understanding of the biochemical strategies regulating reversible poly(ADP-ribosyl)ation. Our results indicate the existence of bacterial PAR metabolism and show that PARG is a member of the widespread macrodomain protein family^{5,6}, which is known to be implicated in genome stability and regulated by poly(ADP-ribosyl)ation^{17–19}. Our data confirm that PARGs are evolutionary unrelated to ARH3 proteins, another class of enzymes with the ability to cleave PAR^{20,21} (Supplementary Information). We believe that our findings will provide the groundwork for future studies that

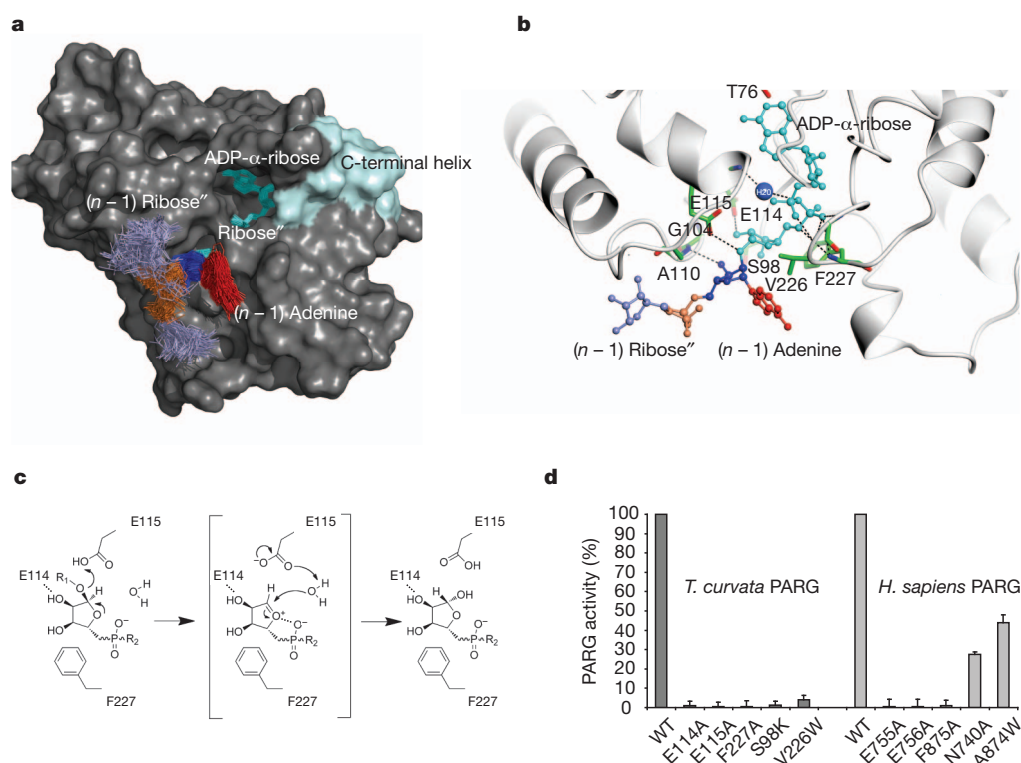


Figure 4 | Structural basis of PAR glycohydrolysis. **a**, The PARG solvent accessible surface derived from molecular dynamics simulations. **b**, A detailed view of the lowest energy PARG-ADP-ribose dimer model obtained from the 100 snapshot structures. Key active site residues are highlighted in green with PAR structural elements coloured as in **a**. **c**, Proposed mechanism for PAR glycohydrolysis. R_1 and R_2 represent $(n - 1)$ PAR and terminal adenosine moieties respectively. **d**, PARG activities of the *T. curvata* PARG mutants and the corresponding *H. sapiens* PARG mutants. WT, wild type. Error bars represent s.d. ($n = 3$).

might ultimately lead to the development of small, cell-permeable PARG inhibitors and the potential to manipulate the physiology of health and disease by interfering with PAR metabolism.

METHODS SUMMARY

Proteins with His₆-tags were purified from *E. coli*. A colourimetric PARG Assay Kit (Trevigen) was used to measure the loss of biotinylated PAR from histones. Auto-modified poly(ADP-ribosyl)ated human and bacterial PARP substrates were made by incubation with NAD¹⁰. *In vivo* activity of PARGs was tested in a yeast system with inducible expression of human PARP1. The *T. curvata* PARG structure was solved by means of single-wavelength anomalous diffraction on Pt-soaked crystals. Ligand-enzyme complexes were obtained by the soaking of PARG crystals. The PAR-PARG model was created using molecular dynamics simulations of the ADP- α -ribose dimer in complex with PARG.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 30 November 2010; accepted 28 July 2011.

Published online 4 September 2011.

- Hakmé, A., Wong, H. K., Dantzer, F. & Schreiber, V. The expanding field of poly(ADP-ribosylation) reactions. *EMBO Rep.* **9**, 1094–1100 (2008).
- D'Amours, D., Desnoyers, S., D'Silva, I. & Poirier, G. G. Poly(ADP-ribosylation) reactions in the regulation of nuclear functions. *Biochem. J.* **342**, 249–268 (1999).
- Koh, D. W. *et al.* Failure to degrade poly(ADP-ribose) causes increased sensitivity to cytotoxicity and early embryonic lethality. *Proc. Natl Acad. Sci. USA* **101**, 17699–17704 (2004).
- Hanai, S. *et al.* Loss of poly(ADP-ribose) glycohydrolase causes progressive neurodegeneration in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **101**, 82–86 (2004).
- Karras, G. I. *et al.* The macro domain is an ADP-ribose binding module. *EMBO J.* **24**, 1911–1920 (2005).
- Till, S. & Ladurner, A. G. Sensing NAD metabolites through macro domains. *Front. Biosci.* **14**, 3246–3258 (2009).
- Patel, C. N., Koh, D. W., Jacobson, M. K. & Oliveira, M. A. Identification of three critical acidic residues of poly(ADP-ribose) glycohydrolase involved in catalysis: determining the PARG catalytic domain. *Biochem. J.* **388**, 493–500 (2005).
- Panda, S., Poirier, G. G. & Kay, S. A. *tej* defines a role for poly(ADP-ribosylation) in establishing period length of the *Arabidopsis* circadian oscillator. *Dev. Cell* **3**, 51–61 (2002).
- Koh, D. W. *et al.* Identification of an inhibitor binding site of poly(ADP-ribose) glycohydrolase. *Biochemistry* **42**, 4855–4863 (2003).
- Ahel, I. *et al.* Poly(ADP-ribose)-binding zinc finger motifs in DNA repair/checkpoint proteins. *Nature* **451**, 81–85 (2008).
- Kothe, G. O., Kitamura, M., Masutani, M., Selker, E. U. & Inoue, H. PARP is involved in replicative aging in *Neurospora crassa*. *Fungal Genet. Biol.* **47**, 297–309 (2010).
- Semighini, C. P., Savoldi, M., Goldman, G. H. & Harris, S. D. Functional characterization of the putative *Aspergillus nidulans* poly(ADP-ribose) polymerase homolog PrpA. *Genetics* **173**, 87–98 (2006).
- Hassa, P. O. & Hottiger, M. O. The diverse biological roles of mammalian PARPs, a small but powerful family of poly-ADP-ribose polymerases. *Front. Biosci.* **13**, 3046–3082 (2008).
- Tao, Z., Gao, P. & Liu, H. W. Studies of the expression of human poly(ADP-ribose) polymerase-1 in *Saccharomyces cerevisiae* and identification of PARP-1 substrates by yeast proteome microarray screening. *Biochemistry* **48**, 11745–11754 (2009).
- Lin, W., Ame, J. C., Aboul-Ela, N., Jacobson, E. L. & Jacobson, M. K. Isolation and characterization of the cDNA encoding bovine poly(ADP-ribose) glycohydrolase. *J. Biol. Chem.* **272**, 11895–11901 (1997).
- Kustatscher, G., Hothorn, M., Pugieux, C., Scheffzek, K. & Ladurner, A. G. Splicing regulates NAD metabolite binding to histone macroH2A. *Nature Struct. Mol. Biol.* **12**, 624–625 (2005).
- Ahel, D. *et al.* Poly(ADP-ribose)-dependent regulation of DNA repair by the chromatin remodeling enzyme ALC1. *Science* **325**, 1240–1243 (2009).
- Gottschalk, A. J. *et al.* Poly(ADP-ribosylation) directs recruitment and activation of an ATP-dependent chromatin remodeler. *Proc. Natl Acad. Sci. USA* **106**, 13770–13774 (2009).
- Timinszky, G. *et al.* A macrodomain-containing histone rearranges chromatin upon sensing PARP1 activation. *Nature Struct. Mol. Biol.* **16**, 923–929 (2009).
- Mueller-Dieckmann, C. *et al.* The structure of human ADP-ribosylhydrolase 3 (ARH3) provides insights into the reversibility of protein ADP-ribosylation. *Proc. Natl Acad. Sci. USA* **103**, 15026–15031 (2006).
- Oka, S., Kato, J. & Moss, J. Identification and characterization of a mammalian 39-kDa poly(ADP-ribose) glycohydrolase. *J. Biol. Chem.* **281**, 705–713 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank G. Clark for genomic DNA from *E. dispar*, G. Smith for the PARP inhibitor, M. Rossi for purified proteins and R. Thorough for editing English. We thank D. Ahel, A. Jordan, D. Ogilvie, S. Terzic, D. Neuhaus and S. Eustermann for helpful discussions. We are grateful to B. Lüscher for the gift of the PARP10 expression plasmid, and K. Labib and the members of his laboratory for advice with yeast work. This work was funded by Cancer Research UK. D.S. holds an AXA Research Fund post-doctoral fellowship. D.L. is a Royal Society University Research Fellow. Access to Diamond beamlines is gratefully acknowledged.

Author Contributions D.S. performed biochemical and *in vivo* experiments, prepared proteins for crystallization, analysed data and wrote the manuscript. M.S.D. performed structural/biophysical studies and analysed data. E.B. performed biochemical and *in vivo* experiments; R.W. performed supporting studies. M.A. and N.D. performed LC/MS analyses; P.L. performed molecular modelling studies. I.A. and D.L. wrote the manuscript, designed experiments and analysed data.

Author Information Atomic coordinates and structure factors have been deposited with the Protein Data Bank under accession codes 3SIG, 3SIH, 3SII and 3SIJ. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to I.A. (iahel@picr.man.ac.uk).

METHODS

Plasmids and proteins. Bacterial and fungal PARG proteins and the *E. dispar* PARG were expressed from the pET28a vector (Novagen). Human PARG with an N-terminal truncation ($\Delta 1-455$) was expressed from the pColdTF vector (Takara). The *A. fumigatus* Af293 gene (AFUA_4G11940), the *D. radiodurans* R1 ATCC13939 gene (DR_B0099) and the *E. dispar* SAW760 gene (EDI_110590) were amplified from genomic DNA, the human *PARP1* and *PARG* genes (Q86W56) were amplified from HeLa cDNA, while the *T. curvata* DSM 43183 gene (Tcur_1721) and the *A. variabilis* ATCC 29413 gene (Ava_4013) were synthesized according to the database sequence (GenScript USA). The *H. aurantiacus* ATCC 23779 *PARG* (Haur_1618) and *PARP* (Haur_4763) genes were amplified from genomic DNA extracted from a dried culture purchased from DSMZ. The codes in brackets indicate gene accession numbers. All proteins bear an N-terminal His tag with the exception of the *A. variabilis* *PARG*, which has a C-terminal His tag. Mutations were introduced using the QuickChange II site-directed mutagenesis kit (Stratagene). Proteins were expressed in *E. coli* Rosetta2(DE3) cells (Novagen). Recombinant proteins were purified on Ni-NTA beads according to standard procedure. For crystallization studies, the *T. curvata* *PARG* was purified by FPLC on a HisTrap HP column followed by Superdex 200 (GE Healthcare). The catalytic fragment of human *PARP1* (amino acids 818–1025) was expressed as a GST-tagged protein and purified on glutathione sepharose columns.

PARP activity assay. PARP automodification activity of the recombinant *H. aurantiacus* PARP was assayed in a 30-min room-temperature reaction containing 100 nM *H. aurantiacus* PARP, 200 μ M NAD (Trevigen), 50 mM Tris pH 7.5 and 50 mM NaCl with or without activated DNA (Trevigen), and with or without the PARP inhibitor KU-0058948 (10 μ M). After the reactions were stopped with the PARP inhibitor, 500 nM *H. aurantiacus* PARG and its catalytic glutamate mutants were added to the reaction for 30 min. The reactions were analysed by western blotting with rabbit anti-PAR antibodies as described later.

PARG activity assays. PARG activity of recombinant proteins was quantified with a colorimetric PARG assay kit (Trevigen), which measured the loss of biotinylated PAR from histones generated by *PARP1*. 3 nM PARG enzymes were used for this assay. For western-blot analysis of PARG activity, PAR was synthesized by the automodification of *PARP1* in a reaction mixture containing 2 units of *PARP1* (Trevigen), 200 μ M NAD (Trevigen), activated DNA (Trevigen), 50 mM Tris pH 7.5 and 50 mM NaCl at 20 °C. Reactions were stopped after 30 min by the addition of the PARP inhibitor KU-0058948. 500 nM PARGs were added to the reaction and incubated for another 30 min. The reactions were run on 4–12% SDS-PAGE gels and blotted onto a nitrocellulose membrane. PAR hydrolysis was detected by rabbit polyclonal anti-PAR antibodies (1:1,000; Trevigen). Western blots were analysed densitometrically by GeneTools (SynGene). For a direct visualization of PARG activity via SDS-PAGE, [32 P]-labelled NAD (GE Healthcare) was added to the above reaction to generate radioactively labelled poly(ADP-ribosyl)ated or mono(ADP-ribosyl)ated PARP substrates. 500 nM PARGs were subsequently added with or without the PARG inhibitor ADP-HPD (50 μ M). The reaction products were analysed by SDS-PAGE and visualized by autoradiography. For TLC analysis, the reactions were stopped by the addition of 50 mM ADP-ribose, spotted onto polyethyleneimine (PEI)-cellulose plates (Macherey-Nagel, Polygram CEL 300 PEI/UV₂₅₄) and developed in 0.15 M LiCl and 0.15 M formic acid. Dried plates were exposed on X-ray film or visualized by UV₂₅₄ shadowing.

Expression of heterologous proteins in *S. cerevisiae*. Plasmids used to introduce human *PARP1*, *A. fumigatus* PARG and *D. radiodurans* PARG into *S. cerevisiae* were created via the Gateway (Invitrogen) site-specific recombination system using yeast-specific destination vectors (Addgene)²². Human *PARP1* was introduced into pAG303GAL-ccdB (HIS3), while *A. fumigatus* PARG and *D. radiodurans* PARG were introduced into pAG423GAL-ccdB (URA3). The host strain W303-1a (MATa, *ade2-1*, *ura3-1*, *his3-11,15*, *trp1-1*, *leu2-3,112*, *can1-100*) was grown in rich YPD medium at 30 °C. The strain was first transformed with the human *PARP1*-harbouring plasmid resulting in His⁺ transformants with chromosomally integrated human *PARP1*. Subsequent transformation with the PARG-harbouring plasmids yielded His⁺Ura⁺ transformants with episomally maintained PARG. To induce the expression of *PARP1* and PARG from the *GAL1* promoter, the strains were grown overnight in minimal SC medium with 2% glucose. The overnight cultures were washed, resuspended in SC+2% raffinose at 7×10^6 cells ml⁻¹ and grown for one cell generation. Cells were centrifuged before resuspending in SC+2% galactose medium and grown overnight to induce the expression of human *PARP1* and PARG from the *GAL1* promoter. Yeast protein extracts were analysed by immunoblotting. Rabbit polyclonal anti-PAR antibody (Trevigen) and rabbit anti-*PARP1* antibody (Abcam) were used at 1:1,000 dilution, while rat monoclonal anti- α -tubulin antibody (Abcam) was used at 1:5,000 dilution. Rabbit anti-*D. radiodurans* PARG and anti-*A. fumigatus* PARG polyclonal antibodies were raised

against the respective full-length proteins purified from *E. coli* (Eurogentec) and applied at 1:1,000 dilution.

Crystallization, refinement and model building. Initial *T. curvata* PARG crystallization conditions were identified using the JCSG+ matrix screen (Molecular dimensions). Crystals suitable for diffraction experiments were obtained by sitting-drop vapour diffusion in 4 μ l drops containing equal volumes of protein and a solution containing 10% PEG 6K and 0.1 M Bicine buffer (pH 9.0). The crystals were derivatised by soaking with the same solution supplemented with 1 mM potassium tetracyanoplatinate (II) trihydrate for 10 min. For the ADP-ribose and ADP-HPD complexes, native crystals were soaked with 1 mM ligand for 1–2 h. The crystals were cryoprotected with the addition of 15% PEG 200 to the mother liquor and flash cooled in liquid nitrogen. Data were collected on beamline I04 at the Diamond Light Source Facility and reduced and scaled with the X-ray Detector Software suite (XDS)²³.

The ligand-free PARG crystal structure was determined by single-wavelength anomalous diffraction on a Pt-soaked crystal using Solve and Resolve as implemented in Phenix²⁴. The initial electron density map was of sufficient quality for Resolve to build 70% of the model. Further automated model building with ARPwARP²⁵ successfully constructed the entire model, with the exception of a small loop region incorporating residues 58–66, for which no corresponding density could be observed. The model was completed by iterative cycles of manual model building and real space refinement using the program Coot²⁶ and crystallographic refinement using Phenix.refine²⁴. Validation of the structure was performed with Molprobit. The processing, phasing and final refinement statistics are presented in Supplementary Table 1.

ITC. The thermodynamic ligand binding properties of wild-type *T. curvata* and two mutant proteins (E114A and E115A) were measured using a VP-ITC microcalorimeter. Protein and ligand concentrations were 20 μ M and 200 μ M respectively in 50 mM Tris (pH 7.5), 100 mM NaCl. Titration curves were fitted using a nonlinear least squares method in Microcal Origin software. A model that was indicative of a single binding site was found to give the best fit in each case; this model was used to obtain the thermodynamic parameters shown in Supplementary Fig. 9.

Thermal shift assay. In each 50 μ l reaction 15 μ l of 300 \times Sypro Orange (Sigma), 2 μ l of water and 33 μ l of protein (4 mg ml⁻¹) in 25 mM Tris pH 7.5, 50 mM NaCl, 10% glycerol and 1 mM DTT were added to the wells of a 96-well thin-wall PCR plate (Bio-Rad). For thermal shift assays in the presence of ADP-ribose, 2 μ l of 25 mM ADP-ribose was added instead of water. Samples were placed in a CFX96 Real Time PCR thermal cycler (Bio-Rad) and slowly heated from 15 to 95 °C with sample fluorescence recorded every 0.2 °C. Fluorescence was monitored at 575 nm (emission), using 490 nm for excitation.

Analysis of PAR chains by sequencing gels. [32 P]-labelled PAR chains produced by PARPs were resolved in 10% sequencing gels in TBE buffer. Reactions were stopped by the addition of 600 μ l 20% ice-cold trichloroacetic acid (TCA). The precipitated material was washed twice with 5% TCA. The pellet was then incubated with 70 μ l of 10 mM Tris and 1 mM EDTA buffer (pH 12) for 2 h at 60 °C to detach PAR chains. PAR was phenol-chloroform extracted, dried and resuspended in 2 mM EDTA-formamide buffer before loading. The products were visualized by autoradiography.

Computational simulations. NAMD software²⁷ was used to perform all molecular dynamics simulations of the ADP- α -ribose dimer in complex with PARG. The topology and parameters files for the ligand were obtained using the Antechamber program²⁸ and AM1-BCC charges²⁹. The additional ADP- α -ribose monomer was added to the ADP- α -ribose bound to PARG by creating an ($\alpha 1' \rightarrow 2'$) O-glycosidic bond. Five conformers of the added monomer were taken as starting points for molecular dynamics simulations. The PARG-ADP- α -ribose dimer complex model was then immersed in a periodic water box (TIP3), as half of the ligand was in contact with the bulk solvent. Then, for each conformer, several simulations (0.5 ns) were run with variable parameters (200 K or 310 K, peptidic backbone fixed, restrained or free). During these MD simulations, snapshots (100 in total for clarity) were periodically extracted and energy-minimized, to represent the conformational flexibility of the ADP- α -ribose dimer.

Analyses by UHPLC coupled to QTOFMS. Ultrahigh-performance liquid chromatography (UHPLC) coupled to quadrupole-time-of-flight mass spectrometry (QTOFMS) was used to analyse the nucleotide products of PARG enzymes. The poly(ADP-ribosyl)ated *PARP1* was treated with *T. curvata* PARG, and the mixture was filtered using centricons (30 kD cut-off). The analysis of the filtrate was performed using a modified previously described procedure³⁰, which used ion-pair chromatography for the separation of nucleotides and negative electrospray ionization for mass spectrometric detection. All analyses were performed using a Waters Acquity UPLC system (Waters Corp.), equipped with a binary solvent delivery system and autosampler. The chromatographic separations used a column (50 mm \times 2.1 mm) filled with a 1.7 μ m BEH C18 stationary phase

(Waters Corp.). Binary gradients at a flow rate of 0.4 ml min^{-1} were applied for the elution. The eluent A was water containing 5 mmol l^{-1} of pentylamine, while the pH value was adjusted to 6.5 using acetic acid. A fast elution gradient was applied, which allowed an efficient separation of NAD^+ and ADP-ribose within 5 min. The gradient started at 2% B and then the percentage of B linearly increased to 25% in 5 min.

The mass spectrometry was performed on a QTOF Premier instrument (Waters Micromass) using an orthogonal Z-spray-electrospray interface. The drying gas and the nebulizing gas was nitrogen. The desolvation gas flow was set to 700 l h^{-1} at a temperature of 300°C . The cone gas flow was adjusted to 25 l h^{-1} , and the source temperature to 120°C . The capillary and cone voltages were 3,200 V and 30 V respectively. The instrument was operated in V mode with TOFMS data being collected between m/z 100–1,000, applying a collision energy of 4 eV. All spectra were recorded using the extended dynamic range (DRE) option in order to correct for possible peak saturations, the data were collected in the centroid mode with a scan time of 0.08 s and interscan time of 0.02 s. In order to ensure maximum accuracy and reproducibility of the system, all acquisitions were carried out using an independent reference spray via the lock spray interface. Leucine enkephalin was applied as a lock mass in negative ionization mode (m/z 554.2615).

The chromatograms, recorded in the total ion current (TIC) mode, were systematically examined by manually generating mass spectra of each visible individual peak using the background-subtraction option and searching for target components expected to occur in the mixture, in particular ADP-ribose and NAD^+ , using extracted ion chromatograms. The target analyses were based on the accurate mass feature of the instrument, applying a mass window of 50 mDa.

Elemental composition of the selected ions was calculated using the MassLynx software incorporated in the instrument. The main criterion for selecting the elements to be included in search parameters for the calculation of the molecular formula was the elemental composition of the presumed parent compounds. According to the instrument specifications, the acceptable deviation from the theoretical m/z values was set at 5 mDa.

22. Alberti, S., Gitler, A. D. & Lindquist, S. A suite of Gateway cloning vectors for high-throughput genetic analysis in *Saccharomyces cerevisiae*. *Yeast* **24**, 913–919 (2007).
23. Kabsch, W. Evaluation of single-crystal X-ray diffraction data from a position-sensitive detector. *J. Appl. Cryst.* **21**, 916–924 (1988).
24. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
25. Perrakis, A., Morris, R. & Lamzin, V. S. Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6**, 458–463 (1999).
26. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
27. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).
28. Cornell, W. D. *et al.* A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117**, 5179–5197 (1995).
29. Jakalian, A., Jack, D. B. & Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **23**, 1623–1641 (2002).
30. Coulrier, L. *et al.* Simultaneous quantitative analysis of metabolites using ion-pair liquid chromatography-electrospray ionization mass spectrometry. *Anal. Chem.* **78**, 6573–6582 (2006).

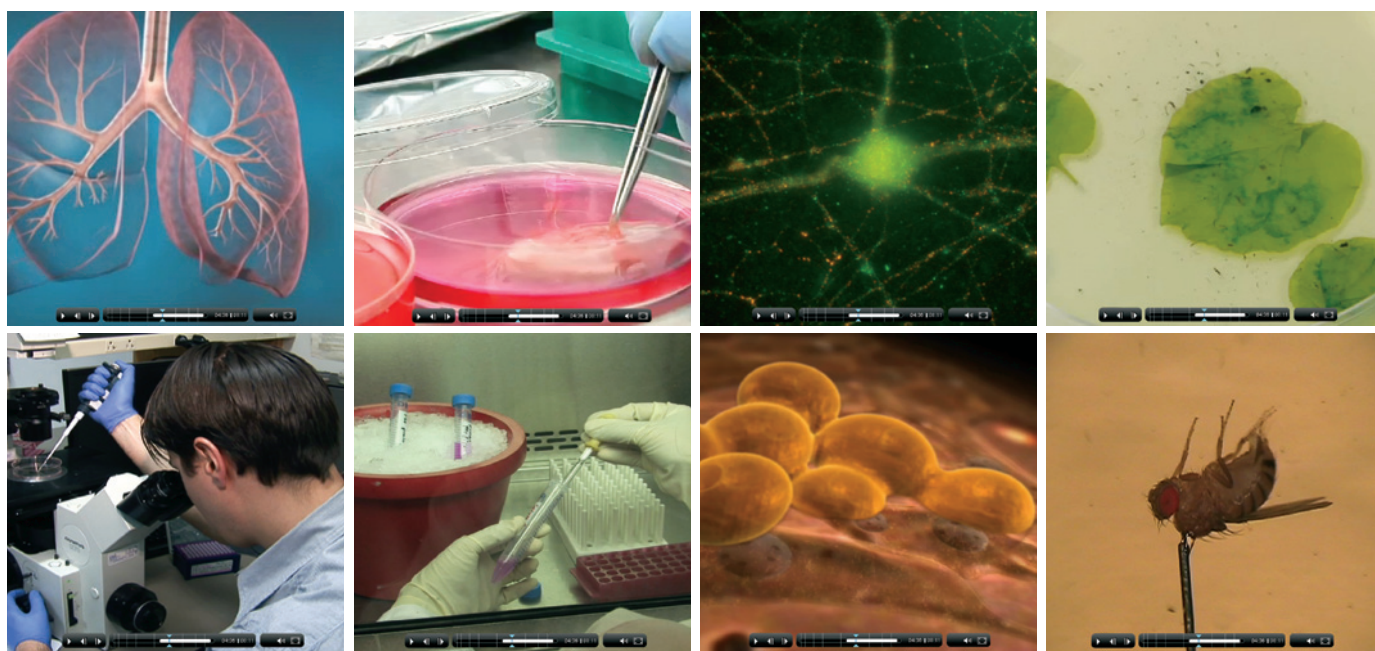
CAREERS

TURNING POINT Pathologist pioneers fetal diagnostic tests **p.623**

ACADEMIA Yale University plugs in new energy-research centre **p.623**

NATUREJOBS For the latest career listings and advice www.naturejobs.com

JOVE



Online videos can show researchers details of a technique that they could never glean from written descriptions in journal articles.

LEARNING TOOLS

Visual aids

Internet-based tools and videos are making it easier to perfect lab techniques and tasks. But they augment, rather than replace, conventional guidance in person.

BY KATHARINE SANDERSON

During her PhD, neuroscientist Maria Toledo-Rodriguez needed to use the single-cell reverse-transcription polymerase chain reaction, a specialized technique that combines electrophysiology with molecular biology. But only a handful of labs in the world had experts in this technique, and she could not manage to visit them. Toledo-Rodriguez, who is now an academic fellow at the University of Nottingham, UK, spent two years struggling to master the technique by following written instructions.

She wishes that she'd been able to take advantage of a now-flourishing type of tutelage: watching a video of someone performing the technique. "There are very stupid things like when you remove the pipette, in order to avoid bursting the neuron, you have to give it a little kick," she says. This is not the kind of

detail that researchers generally put in a methods paper — and Toledo-Rodriguez says that seeing the technique in action gives a much better feel for the nuances than reading a description of it.

Videos that demonstrate how to run experiments, including all those 'stupid' little tricks, are appearing online. They could save time and money for researchers, and allow them to master techniques outside of their subfield.

THE TACIT TOUCH

Mentors and experienced hands aren't always able to offer the time and attention that it takes to teach a 'technique rookie' the nuances and tacit skills that make hands-on science smoother, so videos are becoming a major asset to those hungry for an expanded laboratory skill set. Lab mates and collaborators, with their capacity for relating the feel of a successfully executed technique, can never be replaced

entirely, but film clips go a long way.

Video protocols can be especially helpful when a researcher is switching fields, or delving into new research territory. For example, when Rachel Schecter moved from researching yeast as an undergraduate to examining the neuroscience of autism as a PhD student at the Massachusetts Institute of Technology (MIT) in Cambridge, she had to learn a whole new set of practical techniques. "I found it especially hard when I was starting out because the people in the lab who can teach you these things are often the busiest with their own stuff," says Schecter. "So you really have to maximize the time they have to offer you."

Schecter became a fan of benchfly.com, a website that collects and organizes videos of laboratory techniques. The clips proved useful as tutorials, or as introductions to a method before she worked on it in person with someone from the lab. They cover everything ►

► from basic practical techniques, such as how to work in a sterile hood or use a pipette, to more technical procedures, including making primary neuron cultures, implanting electrodes or running a two-photon microscope. The videos help Schecter to get the “best use of the mentoring in the lab”, she says. And they allow her to double-check her technique after her initial training. “I find reading journal article methods sections incredibly frustrating. They don’t have the detail,” says Schecter, likening the experience to following a cake recipe that says only to mix the ingredients, without describing exactly how much mixing to do or how vigorous it should be.

BenchFly’s videos also offer tips on problems facing lab newcomers, including the most common reason that a centrifuge won’t switch on (a switch that starts the centrifuge only after a delay gets flipped accidentally). The site is organized by discipline and technique. And each video has a comment facility. As on sites such as YouTube, videos can be uploaded by anyone, from academics to students and lab-equipment companies.

Alan Marnett founded BenchFly after a frustrating experience as a chemistry doctoral student at the University of California, San Francisco. “I constantly changed fields,” says Marnett. “I was faced at every step with the challenges of a new environment.” After his PhD, he pursued a postdoc at the Picower Institute for Learning and Memory at MIT, studying how neurons grow. As he tried to bone up on new techniques, he found that he received inconsistent training. “It is only as good as the people around you,” he says. Marnett launched BenchFly in July 2009. The site sustains itself with money earned through advertising and video production services.

Video tutorials save not only time and effort from mentors, but also money and even travel, as Moshe Pritsker realized. As a PhD student working on stem cells at Princeton University in New Jersey, Pritsker was asked to recreate a method of culturing embryonic stem cells that had been reported by researchers in Edinburgh, UK. “I tried to follow the steps in an article,” he says. But try as he might, the experiment wouldn’t work.

His supervisor — keen to have someone in the lab who could reliably do the technique — flew Pritsker to Edinburgh for a two-week reconnaissance and learning mission with the group that had invented it. On the return

flight, Pritsker started to question the necessity of what had just happened — flying across the Atlantic Ocean to watch someone else do an experiment. “This is not the twenty-first century, this is the Stone Age,” Pritsker recalls thinking.

His solution was to “bring the ‘show me’ effect” to scientists through videos of experiments. But to sell the concept, he decided to present the clips in the form of a peer-reviewed journal. “I realized we needed an incentive for scientists to watch and make videos,” he says. The result was the *Journal of Visualized Experiments (JOVE)*, which Pritsker created in 2006 after finishing a postdoc at Harvard Medical School in Boston, Massachusetts.

Whereas BenchFly tends to show basic techniques, or small tricks to get something going, *JOVE* demonstrates detailed, specific and recently developed protocols. And in contrast to BenchFly’s do-it-yourself approach, *JOVE*’s videos are filmed and produced by an international network of trained camera operators and producers. Production takes place after a text account of the video’s technique is peer-reviewed and accepted for publication. Learning the craft and technical aspects of science is what takes up most of a lab’s time and money, says Pritsker — and videos can help to curb such excesses.

They may also be able to facilitate student comprehension. Videos are a great resource for exploring the finer technical points of a topic, says Alice Rushforth, a programme manager at the Center for Emergent Behaviors of Integrated Cellular Systems (EBICS) at MIT. She used BenchFly to teach undergraduates in her previous job as an MIT biology instructor. Rushforth notes that it’s not always possible to show every student every technique individually — it’s both expensive and time consuming, and videos can fill the gap.

Rushforth hopes to use BenchFly to make a package of videos for EBICS summer-school students. Toledo-Rodriguez notes that videos are especially helpful for scientists who teach general university courses involving techniques that aren’t necessarily within their area of expertise.

IN LIVING COLOUR

Of course, videos have their limitations. “There is no better way to learn a new technique than to stand beside an experienced practitioner and watch them work,” says Chris Surridge, editor of *Nature Protocols*. And it’s hard to replicate the benefits of a colleague’s scrutiny and correction. Toledo-Rodriguez has used *JOVE* videos to learn methods for dealing with stem cells, but says that sometimes it is not enough. “The videos go very fast, so for little details such as solution concentrations, you need the numbers, and it’s always better to have the paper,” she says.

And then there’s the possibility that the video protocol will be wrong. Marnett says

that BenchFly does exert quality control, trying to make sure that erroneous techniques don’t slip through. “We will take a video down only if it incorrectly demonstrates a technique or if we feel it’s unnecessarily dangerous,” he says. So far, there



Training in person “is only as good as the people around you”.

Alan Marnett

have been no such cases. But there are often several correct ways of performing any given technique, so BenchFly must be careful not to be too strict about what it allows. “If a video is posted that isn’t exactly how we would do it, but it is scientifically sound, then we would certainly leave it there,” says Marnett.

Videos that accompany peer-reviewed articles are likely to be more accurate than those that are posted with minimal quality control, and some print journals are starting to request more videos. *Nature Protocols* has a section on its website (go.nature.com/bcjwzs), that hosts videos included with the supplementary information accompanying papers — but such submissions remain rare, says Surridge.

The journal also has a YouTube channel (go.nature.com/alxr4h) dedicated to videos from papers or from the protocol exchange, an open-access repository run by *Nature Protocols*. Scientists can use the resource to share their own tips and videos of techniques: they sign up to be part of an online ‘lab group’, then they can post protocols, which must follow a required format and aren’t necessarily videos. The chemistry journal *Angewandte Chemie* also allows researchers to submit video supplements.

But video supplements can be difficult to find, as they’re not usually collated in a central location on journal websites, and they don’t necessarily show experimental techniques; rather, they might demonstrate a colour change in a chemical reaction, or another visual result of the study with which they are associated.

For an increasing number of researchers, however, videos provide a way to get up to speed in a new field and can supplement the unpredictable technical know-how of lab mates and supervisors. “Becoming a great scientist,” says Marnett, “is just as much about learning the tips and tricks that increase quality and reproducibility of data as it is about learning the science behind the experiments.” ■

Katharine Sanderson is a freelance writer based in Toulouse, France.



“I find reading methods sections incredibly frustrating. They don’t have the detail.”

Rachel Schecter

WOMANSPACE

Parallel processing.

BY ED RYBICKI

You'd never think that all it took was two middle-aged men, sent shopping by the wife of one to buy knickers, to crack the biggest discovery in modern physics.

It was very simple: I'd been staying with my friend Russell in Canberra, trying to sort out how we were going to get our book on virus structure together, when Russell's wife Lilia decided that their youngest daughter needed new school knickers. She was too busy making supper to bother; these otherwise unemployed elderly men were the perfect candidates — and the prospect of not having to listen to us blather on about just where to pitch the book, and what to put in it, and which Jethro Tull albums we liked, probably tipped the balance our way. Seeing as we could continue to do all those things in a car and in the supermarket — and do a side trip to drool over new electronic goodies in Harvey Norman — we agreed with alacrity.

And so it was, that after a most satisfying comparison shop of iMacs versus the rest, and a cruise through rock nostalgia in the shape of special-offer CDs, we found ourselves in a large supermarket, trawling for girls' knickers.

At this point I must digress, and mention, for those who are not aware, the profound differences in strategy between Men Going Shopping and Women Going Shopping. In any general shopping situation, men *hunt*: that is, they go into a complex environment with a few clear objectives, achieve those, and leave. Women, on the other hand, *gather*: such that any mission to buy just bread and milk could turn into an extended foraging expedition that also snares a to-die-for pair of discounted shoes; a useful new mop; three sorts of new cook-in sauces; and possibly a selection of frozen fish.

And the interesting thing is — and this is what sparked the discovery — that any male would be very hard pressed to say where she got some of these things, even if he accompanied her.

Have you never had the experience of talking to your significant female other as you wend your way through the complexity of a supermarket — only to suddenly find her 20 metres away with her back to you? And then she comes back with something



you've never seen before, and tosses it in the trolley as if nothing has happened?

I know I have — and until recently, I had always assumed it was just me not noticing new things in the aisles we were walking through.

So there we were, looking for knickers, and a rather wary woman asked if she could help, given that we looked lost and hopeless. Russell explained to her exactly what we were looking for, and her wariness seemed to become mild alarm, until we hastened to reassure her that this was in fact a commission for the mother of said child. She then said, with what seemed to be great satisfaction, "Oh, no; you'll never find those in here — you'll have to go down to [some remote location]," which we had no chance of achieving before they closed, so the whole mission was now a failure.

It was as we trudged our forlorn way back to the car that Russell said: "You know, I'm sure we've found them there before — at least, Lilia has."

I said, only half-joking: "Well, women seem to be able to do that — maybe they're getting into spaces we poor guys can't?"

That was the catalyst: suddenly, we Hunters had an insight into how real Gatherers operated, sparked by our own hopelessness and some considerable acquaintance with the formidable talents of wives when it comes to finding things, and enough science (and science fiction)

➤ NATURE.COM
Follow Futures on
Facebook at:
go.nature.com/mtoodm

background to be able to appreciate that parallel universes were quite a reasonable answer to a

number of important questions. Including, it seemed, supermarket shopping.

Oh, we were roundly scorned when we got back, knickerless; both for the lack of same, and for our absurd idea — which smacked of desperate snatching at straws to excuse incompetence, to the astrophysics-qualified wife. However, we persevered: we talked it through exhaustively, even conquering problems like:

"What happens when you take some item from another universe to the checkout — when they won't recognize the barcode?" Except they do — electronic systems are quantum-computing devices in their ability to access stock codes, and when they don't, then it's invariably a woman who gets sent to look for them, and they, of course, find them ...

And there it might have sat, had it not been for the Internet. We simply put the idea up in as many forums as we could access; we blogged on it; we talked to everyone we knew (well, male, obviously) who could be relied on to observe such phenomena — and slowly, the observations came in.

Then, of course, there was arguing about significance, and power of the statistical methods used, and it all got usurped by some theoretical physicists the moment it started to look as though there was something in it.

But the answer is clear: women can access parallel universes in order to find things, whether they do it consciously or not. They have probably always been able to do this, and now there is fierce speculation as to whether this constituted the evolutionary advantage we had over other primates: the presence of bulbs, grains and nuts on the table that had been retrieved from parallel universes when the hunters came home empty-handed was probably a major factor in the survival of our species.

The difference is that now they know that they can do it — and things have changed.

Because groceries aren't all they go looking for. It turns out the next item on the shopping list is better-looking versions of us.

Russell and I no longer communicate. And we're very lonely. ■

Ed Rybicki swears that at least some of this is a true story. The part about Russell Kightley and the knickers, anyway.

JACEY

Detectable radio flares following gravitational waves from mergers of binary neutron stars

Ehud Nakar¹ & Tsvi Piran²

Mergers of neutron-star/neutron-star binaries are strong sources of gravitational waves^{1–3}. They can also launch subrelativistic and mildly relativistic outflows^{4–8} and are often assumed to be the sources of short γ -ray bursts⁹. An electromagnetic signature that persisted for weeks to months after the event would strengthen any future claim of a detection of gravitational waves¹⁰. Here we present results of calculations showing that the interaction of mildly relativistic outflows with the surrounding medium produces radio flares with peak emission at 1.4 gigahertz that persist at detectable (submillijansky) levels for weeks, out to a redshift of 0.1. Slower subrelativistic outflows produce flares detectable for years at 150 megahertz, as well as at 1.4 gigahertz, from slightly shorter distances. The radio transient RT 19870422 (ref. 11) has the properties predicted by our model, and its most probable origin is the merger of a compact neutron-star/neutron-star binary. The lack of radio detections usually associated with short γ -ray bursts does not constrain the radio transients that we discuss here (from mildly relativistic and subrelativistic outflows) because short γ -ray burst redshifts are typically >0.1 and the appropriate timescales (longer than weeks) have not been sampled.

Gravitational-wave detectors, and in particular the advanced LIGO and Virgo interferometers, are being constructed now with the goal of detecting gravitational waves from binary neutron-star coalescence at distances up to a few hundred megaparsecs (redshift $z \approx 0.1$)¹². The detection of an accompanying electromagnetic signal would complement these efforts, providing an independent confirmation of the discovery and increasing the detectors' effective sensitivity. The search for such an electromagnetic signal has therefore attracted much interest. The radioactive decay of ejected debris from the merger would drive a short-lived supernova-like event¹³. For example, ejection of 0.01 solar masses ($0.01M_{\odot}$) from a merger at a distance of 300 Mpc would result in a faint optical flare that peaks after ~ 1 day (ref. 14). Finding, and especially identifying, such rare and faint events in the crowded variable optical sky is an extremely challenging task. Other authors have speculated on the production of low-frequency radio signals from the interaction of the neutron stars' magnetic fields^{15–17}. These attempts focused on electromagnetic signals that are contemporaneous with, or follow quickly, the merger and the gravitational waves. Unfortunately, these predictions are highly uncertain.

Here we predict a robust radio signal that peaks several weeks after the merger. Numerical simulations show that compact binary mergers launch energetic subrelativistic and mildly relativistic outflows^{4–8}. Ejection sources include unbound tidal tails, and winds driven by neutrino heating, nucleosynthesis and electromagnetic processes^{18–21}, emerging from the proto-neutron star or from an accretion disk. Overall, almost all merger models find a significant ejection of mass and energy. In binary neutron-star mergers, an ejection of about 10^{50} erg at $(0.1–0.2)c$ (where c is the speed of light) and about 10^{49} erg as faster ejecta is a fairly robust prediction. The outflow from black-hole/neutron-star mergers is less certain, but it is possibly more energetic and faster¹⁸ ($\sim 10^{52}$ erg at $0.5c$).

The interaction of this outflow with the surrounding tenuous matter generates a blast wave. Although the outflow may be highly non-uniform

initially, it becomes spherical rather quickly. We therefore consider a spherical outflow with energy E and an initial velocity $c\beta_i$ that propagates into a medium with a constant density, n . If the outflow is not ultrarelativistic, it propagates at a constant velocity until time t_{dec} , when, at a radius R_{dec} , it collects a mass comparable to its own. Time t_{dec} (in days) is given by

$$t_{\text{dec}} = \frac{R_{\text{dec}}}{c\beta_i} \approx 30 E_{49}^{1/3} n_0^{-1/3} \beta_i^{-5/3} \quad (1)$$

Here and in the following, unless stated otherwise, q_x (where q is any parameter) denotes the value of $q/10^x$ in c.g.s. units. At a radius $R > R_{\text{dec}}$, the flow decelerates, assuming a Sedov–Taylor blast wave: $\beta \approx \beta_i (R/R_{\text{dec}})^{-3/2}$.

The blast wave generates magnetic fields and accelerates particles that emit synchrotron radiation. The same microphysics used successfully to model radio emission of type Ibc supernovae^{22,23}, where $\beta \approx 0.2$, and to model late radio emission of γ -ray bursts^{24,25}, where the flow is mildly relativistic, is applicable here. In both cases, the electrons and the magnetic field are found to carry significant fractions of the total internal energy of the shocked gas, $\epsilon_e \approx \epsilon_B \approx 0.1$. The observed spectra reveal a power-law distribution of the electrons' Lorentz factor, γ : $dN/d\gamma \propto \gamma^{-p}$ for $\gamma > \gamma_m = [(p-2)/(p-1)](m_p/m_e)\epsilon_e\beta^2$, where m_p and m_e are the proton and electron masses, respectively, γ_m is the minimal Lorentz factor of the electron's distribution and $p \approx 2–3$.

The radio spectrum is determined by v_m , the synchrotron frequency of electrons with Lorentz factor γ_m , and by v_a , the synchrotron self-absorption frequency (Supplementary Information). The specific flux, F_ν , at a given frequency is strongly suppressed below v_a , and it decreases as $\nu^{-(p-1)/2}$ above v_m and v_a . The signal across the whole spectrum increases at $t < t_{\text{dec}}$. Its behaviour after t_{dec} depends on the relation between the observed frequency, ν_{obs} , and v_m and v_a . The signal peaks at t_{dec} if $\nu_{\text{obs}} > v_{a,\text{dec}}, v_{m,\text{dec}}$ where $v_{a,\text{dec}} \equiv v_a(t_{\text{dec}})$ and $v_{m,\text{dec}} \equiv v_m(t_{\text{dec}})$. Otherwise the signal peaks when $\nu_{\text{obs}} = v_m$ or when $\nu_{\text{obs}} = v_a$, whichever is latest.

The flare characteristics are most sensitive to the initial velocity of the outflow, β_i . Because the brightest radio emission is observed at t_{dec} , a lower value of β_i implies a longer rise time of the radio emission after the merger. Additionally, $v_{m,\text{dec}}, v_{a,\text{dec}}$ and the peak flux at any observed frequency depend strongly on β_i . As mergers are expected to eject an outflow over a range of velocities, we discuss separately below the observed signature of mildly relativistic ($\beta_i \approx 1$) and subrelativistic ($\beta_i \approx 0.2$) ejecta.

A mildly relativistic blast wave with canonical parameters produces a synchrotron spectrum with $v_{a,\text{dec}} \leq v_{m,\text{dec}} \approx 1$ GHz. The strongest signal is then expected at time t_{dec} (a few weeks after the merger) and around 1.4 GHz (Supplementary Information): the peak of the observed specific flux at ν_{obs} in units of millijanskys is

$$F_{\nu_{\text{obs}},\text{peak}}[\nu_{\text{obs}} > v_{m,\text{dec}}, v_{a,\text{dec}}] \approx 0.3 E_{49} n_0^{\frac{p+1}{4}} \epsilon_{B,-1}^{\frac{p+1}{4}} \epsilon_{e,-1}^{p-1} \beta_i^{\frac{5p-7}{2}} d_{27}^{-2} \left(\frac{\nu_{\text{obs}}}{1.4} \right)^{-\frac{p-1}{2}} \quad (2)$$

¹Raymond and Beverly Sackler School of Physics & Astronomy, Tel Aviv University, Tel Aviv 69978, Israel. ²Racah Institute of Physics, The Hebrew University, Jerusalem 91904, Israel.

where d is the distance to the merger (here v_{obs} is in GHz). The peak flux at lower frequencies (<1 GHz) is significantly lower and it is observed at a later time. If the outflow is subrelativistic ($\beta_i \approx 0.1$ – 0.2), then $v_{a,\text{dec}}, v_{m,\text{dec}} \leq 150$ MHz and equation (2) is applicable also in the frequency range of low-frequency radio detectors. The flux peaks at t_{dec} , which is of the order of years, and it is brighter at 150 MHz than at 1 GHz by about an order of magnitude. Note that over the whole expected range of blast wave parameters, $v_a \leq 1$ GHz at all times and the spectrum above 1 GHz is optically thin during the entire evolution. We stress that in radio supernovae, the surrounding dense winds lead at early time to an optically thick spectrum at $v_{\text{obs}} > 1$ GHz, and the transition $v_a = v_{\text{obs}}$ determines the time and flux at the peak. As discussed below, this different spectral signature enables us to distinguish between merger flares and radio supernovae.

The circum-merger density also strongly affects the flare signature. For example, if the surrounding particle density is $\sim 10^{-3} \text{ cm}^{-3}$, the peak flux from a mildly relativistic ejecta decreases to the microjansky level at a distance of 300 Mpc, and the timescale increases by a factor of ten, to a year. A merger taking place in such a density can be detected only up to distances of ~ 100 Mpc (Table 1.) The density is expected to vary significantly, from $n \approx 1 \text{ cm}^{-3}$, in galactic disks, to $n \approx 10^{-6} \text{ cm}^{-3}$, for mergers taking place outside their host galaxies. Because all observed Galactic neutron-star binaries reside within the Galactic disk, where the average density is $n \approx 1 \text{ cm}^{-3}$, a significant fraction of the cosmological mergers are expected also to take place in rather dense environments. We therefore use $n = 1 \text{ cm}^{-3}$ as the canonical density value. If mergers produce short γ -ray bursts, then observations of their afterglows support this value (Supplementary Information).

An intriguing possibility is that compact merger events also eject ultrarelativistic jets that produce short γ -ray bursts⁹ (SGRBs). It is important to examine the relationship between SGRBs and the radio flares discussed here, assuming that mergers are producing SGRBs. An SGRB beamed towards us will be observed in coincidence with the gravitational-wave signal, providing a clear electromagnetic counterpart. Even if the SGRB itself is missed, owing to partial sky coverage, its afterglow will be easily detectable. However, SGRBs are expected to be beamed, and only rarely will one point towards us. A beamed SGRB observed off-axis produces, once it has slowed down, a long-lasting radio ‘orphan’ afterglow²⁶, similar in its characteristics to the mildly relativistic signal discussed above. However, the total energy (corrected for beaming) in the ultrarelativistic jet is at most comparable to—and probably lower than—that of the mildly relativistic ejecta. Consequently the latter will dominate the radio emission.

The radio remnant signals that we consider here, which are generated by subrelativistic and mildly relativistic outflows, could not have been detected in the radio afterglow searches that were carried out following SGRB triggers. The reasons are twofold. First, SGRBs are typically detected at distances of 1–3 Gpc, far beyond the detection horizon of gravitational-wave detectors. Hence the signals are much weaker than those associated with detected gravitational-wave events. Second, SGRB afterglow searches are optimized to detect the emission from ultrarelativistic ejecta pointing towards the observer. Such emission

peaks at higher frequencies and on shorter timescales than the emission from mildly and subrelativistic ejecta discussed here. SGRB afterglow searches are done at a sensitivity of ~ 0.1 mJy at 4.8–8.5 GHz during the first week or two after the bursts (see, for example, refs 27–29). Equation (2) implies that over the distance range 1–3 Gpc, the radio signal of mildly relativistic ejecta with energy 10^{50} erg that propagate into a medium of density $n = 1 \text{ cm}^{-3}$ peaks after ~ 60 days at a flux of ~ 0.01 – 0.1 mJy at 5 GHz. The flux before the peak rises as $(t/t_{\text{dec}})^3$ (Supplementary Information). Therefore, these early radio afterglow searches could not have detected the radio signal, even if the mildly relativistic ejecta had an energy of 10^{51} erg. Thus, the paucity of detected SGRB radio afterglows has no direct implication for the nature of the remnants we discuss here.

A new wave of radio detectors is now coming online. The most sensitive operate at frequencies of 1.4 GHz and higher. Table 1 summarizes the relevant properties of these facilities and their detection horizons. The best facility for a targeted search, following a detection of a candidate gravitational-wave source, is clearly the EVLA. A deep, ~ 50 μ Jy, localized EVLA search of the 10–100 deg² error box of a gravitational-wave trigger³⁰ can detect mildly relativistic ejecta (with an energy of even $\sim 10^{48}$ erg) out to the horizon of advanced gravitational-wave detectors. The upcoming lower-frequency LOFAR sensor array will be more effective in searches for subrelativistic outflows, whose signals peak at LOFAR’s frequencies, thus compensating for LOFAR’s lower sensitivity. LOFAR is also relatively more effective when searching for flares in a low-density medium.

Even before the completion of the advanced gravitational-wave detectors, blind searches can identify radio flares from compact binary mergers. Identification of radio emission from any merger type (binary neutron star or black hole/neutron star) would determine the merger rate, which is a parameter of utmost importance for the design and operation of advanced detectors. With a merger rate of $300 \text{ Gpc}^{-3} \text{ yr}^{-1}$, we expect these facilities to detect ~ 20 remnants from mildly relativistic outflows with an energy of $E = 10^{49}$ erg (and $\sim 1,000$ remnants if $E = 10^{50}$ erg), in a single whole-sky snapshot. LOFAR may detect a dozen transients in a whole-sky survey, even if only a subrelativistic outflow with an energy of 10^{50} erg is ejected (see Supplementary Information for details, as well as for a discussion of ways to distinguish these flares from other possible radio transients).

Remarkably, the observed 5 GHz transient RT 19870422 (ref. 11) shows all the expected properties of the radio remnant of a compact binary merger. At 1 Gpc distance and with a duration of two months, this transient is what we would expect from a mildly relativistic outflow with an energy of $\sim 10^{50}$ erg. The inferred rate of similar transients¹¹, 80 – $20,000 \text{ Gpc}^{-3} \text{ yr}^{-1}$, is fully consistent with the estimates of compact binary mergers. This transient is therefore an excellent candidate to be the first observed radio remnant of a merger. Unfortunately, we cannot rule out the possibility that this is an especially bright radio supernova¹¹. We note, however, that this latter interpretation requires a supernova brighter by an order of magnitude than any radio supernovae previously observed. Simultaneous optical observations or multiwavelength radio observations could have easily distinguished

Table 1 | Observing radio flares

Radio facility	Observing frequency (GHz)	Field of view (deg ²)	One-hour r.m.s.* (μJy)	One-hour detection horizon†			Ten-hour detection horizon‡
				$\beta_i \approx 1$, $E_{49} = 1$, $n_0 = 1$	$\beta_i \approx 1$, $E_{49} = 10$, $n_0 = 1$	$\beta_i = 0.2$, $E_{49} = 10$, $n_0 = 1$, $p = 2.5$	$\beta_i \approx 1$, $E_{49} = 1$, $n_0 = 10^{-3}$, $p = 2$
EVLA	1.4	0.25	7	1 Gpc	3.3 Gpc	370 Mpc	140 Mpc
ASKAP	1.4	30	30	500 Mpc	1.6 Gpc	180 Mpc	70 Mpc
MeerKAT	1.4	1.5	35	500 Mpc	1.6 Gpc	165 Mpc	65 Mpc
Apertif	1.4	8	50	400 Mpc	1.25 Gpc	140 Mpc	50 Mpc
LOFAR	0.15	20	1,000	35 Mpc	90 Mpc	70 Mpc	20 Mpc

Shown are properties and detection horizons (neglecting cosmological corrections) for an observation at different radio facilities of blast waves with various values of β_i , E_{49} , n_0 and p (in all cases, $v_a \approx v_b \approx 0.1$; see text for definitions of these symbols). Information on facilities is available as follows: EVLA (<http://www.aoc.nrao.edu/evla/>); ASKAP (<http://www.atnf.csiro.au/projects/askap/technology.html>); MeerKAT (<http://www.ska.ac.za/meerkat/>); Apertif (<http://www.astron.nl/general/apertif/apertif>); and LOFAR (<http://lofar.org>).

* The root mean squared value of the background noise for one hour of observation.

† The distance at which the observed peak flux is four times the one-hour r.m.s.

‡ The distance at which the observed peak flux is four times the root mean squared value of the background noise for ten hours of observation.

between the two possibilities. Unfortunately, no such observations are available. However, the detection rate implied by this event is very high, indicating that similar events could easily be detected by a relatively small-scale survey and that their nature should be easily probed.

Received 11 February; accepted 11 July 2011.

Published online 28 September 2011.

1. Taylor, J. H. & Weisberg, J. M. A new test of general relativity — gravitational radiation and the binary pulsar PSR 1913+16. *Astrophys. J.* **253**, 908–920 (1982).
2. Esposito, L. W. & Harrison, E. R. Properties of the Hulse-Taylor binary pulsar system. *Astrophys. J.* **196**, L1–L2 (1975).
3. Wagoner, R. V. Test for the existence of gravitational radiation. *Astrophys. J.* **196**, L63–L65 (1975).
4. Rosswog, S., Davies, M. B., Thielemann, F. & Piran, T. Merging neutron stars: asymmetric systems. *Astron. Astrophys.* **360**, 171–184 (2000).
5. Ruffert, M. & Janka, H. Coalescing neutron stars — a step towards physical models. III. Improved numerics and different neutron star masses and spins. *Astron. Astrophys.* **380**, 544–577 (2001).
6. Yamamoto, T., Shibata, M. & Taniguchi, K. Simulating coalescing compact binaries by a new code (SACRA). *Phys. Rev. D* **78**, 064054 (2008).
7. Rezzolla, L., Baiotti, L., Giacomazzo, B., Link, D. & Font, J. A. Accurate evolutions of unequal-mass neutron-star binaries: properties of the torus and short GRB engines. *Class. Quantum Grav.* **27**, 114105 (2010).
8. Kiuchi, K., Sekiguchi, Y., Shibata, M. & Taniguchi, K. Exploring binary-neutron-star-merger scenario of short-gamma-ray bursts by gravitational-wave observation. *Phys. Rev. Lett.* **104**, 141101 (2010).
9. Eichler, D., Livio, M., Piran, T. & Schramm, D. N. Nucleosynthesis, neutrino bursts and gamma-rays from coalescing neutron stars. *Nature* **340**, 126–128 (1989).
10. Kochanek, C. S. & Piran, T. Gravitational waves and gamma-ray bursts. *Astrophys. J.* **417**, L17–L20 (1993).
11. Bower, G. C. *et al.* Submillijansky transients in archival radio observations. *Astrophys. J.* **666**, 346–360 (2007).
12. Smith, J. R. & LIGO Scientific Collaboration. The path to the enhanced and advanced LIGO gravitational-wave detectors. *Class. Quantum Grav.* **26**, 114013 (2009).
13. Li, L. & Paczyński, B. Transient events from neutron star mergers. *Astrophys. J.* **507**, L59–L62 (1998).
14. Metzger, B. D. *et al.* Electromagnetic counterparts of compact object mergers powered by the radioactive decay of r-process nuclei. *Mon. Not. R. Astron. Soc.* **406**, 2650–2662 (2010).
15. Hansen, B. M. S. & Lyutikov, M. Radio and X-ray signatures of merging neutron stars. *Mon. Not. R. Astron. Soc.* **322**, 695–701 (2001).
16. Moortgat, J. & Kuijpers, J. Gravitational waves in magnetized relativistic plasmas. *Phys. Rev. D* **70**, 023001 (2004).
17. Pshirkov, M. S. & Postnov, K. A. Radio precursors to neutron star binary mergings. *Astrophys. Space Sci.* **330**, 13–18 (2010).
18. Rosswog, S. Mergers of neutron star-black hole binaries with small mass ratios: nucleosynthesis, gamma-ray bursts, and electromagnetic transients. *Astrophys. J.* **634**, 1202–1213 (2005).
19. Levinson, A. General relativistic, neutrino-assisted magnetohydrodynamic winds-theory and application to gamma-ray bursts. I. Schwarzschild geometry. *Astrophys. J.* **648**, 510–522 (2006).
20. Metzger, B. D., Piro, A. L. & Quataert, E. Time-dependent models of accretion discs formed from compact object mergers. *Mon. Not. R. Astron. Soc.* **390**, 781–797 (2008).
21. Dessart, L., Ott, C. D., Burrows, A., Rosswog, S. & Livne, E. Neutrino signatures and the neutrino-driven wind in binary neutron star mergers. *Astrophys. J.* **690**, 1681–1705 (2009).
22. Chevalier, R. A. Synchrotron self-absorption in radio supernovae. *Astrophys. J.* **499**, 810–819 (1998).
23. Chevalier, R. A. & Fransson, C. Circumstellar emission from Type Ib and Ic supernovae. *Astrophys. J.* **651**, 381–391 (2006).
24. Frail, D. A., Waxman, E. & Kulkarni, S. R. A 450 day light curve of the radio afterglow of GRB 970508: fireball calorimetry. *Astrophys. J.* **537**, 191–204 (2000).
25. Frail, D. A. *et al.* Accurate calorimetry of GRB 030329. *Astrophys. J.* **619**, 994–998 (2005).
26. Rhoads, J. E. How to tell a jet from a balloon: a proposed test for beaming in gamma-ray bursts. *Astrophys. J.* **487**, L1–L4 (1997).
27. Soderberg, A. M. & Frail, D. A. GRB060313: radio observations. *GCN Circ.* **4884** (2006).
28. van der Horst, A. J. GRB 060801: WSRT radio observations. *GCN Circ.* **5390** (2006).
29. Rhoads, M. H., Chandra, P. & Frail, D. A. ATCA radio observations of short hard burst GRB 071112B. *GCN Circ.* **7095** (2007).
30. Wen, L. & Chen, Y. Geometrical expression for the angular resolution of a network of gravitational-wave detectors. *Phys. Rev. D* **81**, 082001 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Frail, S. Kulkarni, A. Levinson, A. MacFadyen, E. Ofek and S. Rosswog for discussions. This research was supported by an ERC advanced research grant, by the Israeli Center for Excellence for High Energy Astrophysics, by the ISF (grant No. 174/08) and by an IRG grant.

Author Contributions Both authors contributed equally.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.P. (tsvi@phys.huji.ac.il) or E.N. (udini@wise.tau.ac.il).

CORRIGENDUM

doi:10.1038/nature10420

Ephrin Bs are essential components of the Reelin pathway to regulate neuronal migration

Aycan Sentürk, Sylvia Pfennig, Alexander Weiss, Katja Burk & Amparo Acker-Palmer

Nature **472**, 356–360 (2011)

In this Letter we made errors in representative image choice, including mislabelling of images or choosing an image from the inappropriate genotype. In all cases, choice of images was completely independent of the data analysis and so none of the conclusions in our original Letter are affected. We apologise for any confusion these errors may have caused.

Figure 1a depicts a Tbr1 staining of the adult mouse cortex for four different genotypes. In the process of choosing representative pictures that reflect the results of our analysis shown in Fig. 1b, cropped images from original pictures were inadvertently mislabelled and used incorrectly. We provide below a corrected version of Fig. 1a with new representative images for the following genotypes: WT and *Reln*^{+/+}; *Efnb3*^{-/-}. A new high-magnification picture for WT is also shown in the two rightmost panels. Original images for every genotype and additional examples are shown in the Supplementary Information of this Corrigendum.

Figure 1c depicts a Brn1 staining of the E17.5 mouse cortex for five different genotypes. In the process of figure assembly cropped images from original pictures were inadvertently mislabelled and used

incorrectly. We provide below a corrected Fig. 1c with a new image for *Reln*^{+/+}; *Efnb3*^{-/-}. In the ephrinB3 compound mice (*Reln*^{+/+}; *Efnb3*^{-/-}) Brn1⁺ cells aberrantly accumulate in the lower layers of the cortex and do not migrate to the upper layers, resembling the Reeler (*Reln*^{-/-}) phenotype. Original pictures and additional examples are shown in the Supplementary Information of this Corrigendum, where arrows indicate the distribution of Brn1⁺ cells. We have also included results from a new, reproduced experiment recently performed with an additional cohort of animals that shows exactly the same results.

In Fig. 1d, the second panel, labelled '*Reln*^{+/+}; *Efnb3*^{-/-}' should instead be labelled '*Reln*^{+/+}'. In the Methods summary section 'Stimulation of neurons', "Cortical neurons from E14.5 were grown" should instead read "Cortical neurons from E15.5 were grown....".

Further errors in the Supplementary Information of the original Letter are described and corrected in the Supplementary Information of this Corrigendum.

Supplementary Information is linked to the online version of the Corrigendum at www.nature.com/nature.

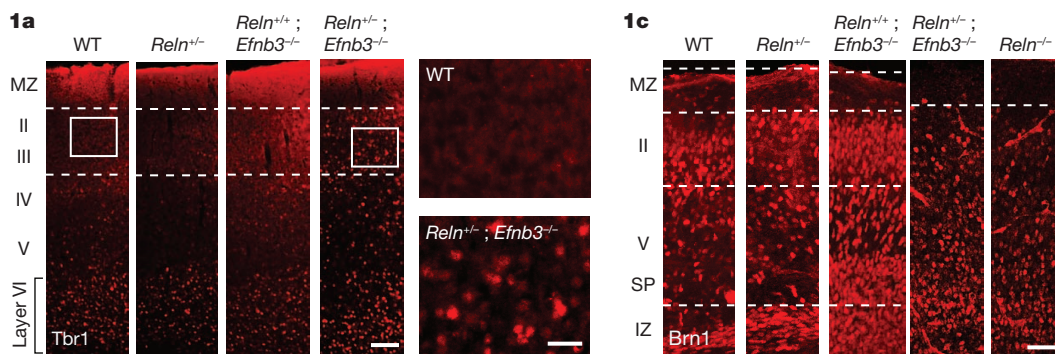


Figure 1 | Corrected Fig. 1a (with two new panels) and corrected Fig. 1c.

Dynamics of human adipose lipid turnover in health and metabolic disease

Peter Arner¹, Samuel Bernard², Mehran Salehpour³, Göran Possnert³, Jakob Liebl⁴, Peter Steier⁴, Bruce A. Buchholz⁵, Mats Eriksson¹, Erik Arner¹, Hans Hauner⁶, Thomas Skurk⁶, Mikael Rydén¹, Keith N. Frayn⁷ & Kirsty L. Spalding⁸

Adipose tissue mass is determined by the storage and removal of triglycerides in adipocytes¹. Little is known, however, about adipose lipid turnover in humans in health and pathology. To study this *in vivo*, here we determined lipid age by measuring ¹⁴C derived from above ground nuclear bomb tests in adipocyte lipids. We report that during the average ten-year lifespan of human adipocytes, triglycerides are renewed six times. Lipid age is independent of adipocyte size, is very stable across a wide range of adult ages and does not differ between genders. Adipocyte lipid turnover, however, is strongly related to conditions with disturbed lipid metabolism. In obesity, triglyceride removal rate (lipolysis followed by oxidation) is decreased and the amount of triglycerides stored each year is increased. In contrast, both lipid removal and storage rates are decreased in non-obese patients diagnosed with the most common hereditary form of dyslipidaemia, familial combined hyperlipidaemia. Lipid removal rate is positively correlated with the capacity of adipocytes to break down triglycerides, as assessed through lipolysis, and is inversely related to insulin resistance. Our data support a mechanism in which adipocyte lipid storage and removal have different roles in health and pathology. High storage but low triglyceride removal promotes fat tissue accumulation and obesity. Reduction of both triglyceride storage and removal decreases lipid shunting through adipose tissue and thus promotes dyslipidaemia. We identify adipocyte lipid turnover as a novel target for prevention and treatment of metabolic disease.

A major function of adipose tissue is to store and release fatty acids, which are incorporated into adipocyte triglycerides according to whole-body energy demands. Body fat mass is determined by the balance between triglyceride storage and removal in adipocytes, by either enzymatic hydrolysis (lipolysis) and subsequent fatty acid oxidation and/or ectopic deposition in non-adipose tissues. Little is known about the dynamics of these processes in humans. Although isotope tracer methods have been used to estimate lipid turnover in human adipose tissue, these studies have been limited to short-term experimental conditions^{1–4}. To study long-term adipose tissue lipid turnover *in vivo* and across the adult lifespan, we developed a method to retrospectively determine the age of adipocyte triglycerides in humans. Triglycerides are the major component of the adipocyte lipid droplet. Lipid age was assessed by measuring the ¹⁴C content in the lipid compartment of adipocytes from human subcutaneous adipose tissue, the major fat depot in humans. ¹⁴C levels in the atmosphere remained remarkably stable until above ground nuclear bomb tests between approximately 1955 and 1963 caused a significant increase in ¹⁴C relative to stable carbon isotope levels⁵ (Fig. 1a). After the Limited Nuclear Test Ban Treaty was signed in 1963, ¹⁴C levels in the atmosphere decreased exponentially. This is not due to radioactive decay (half-life ($T_{1/2}$) for ¹⁴C is 5,730 years), but to diffusion of ¹⁴CO₂ out of the atmosphere⁶. ¹⁴C in the atmosphere oxidises to form CO₂, which is

taken up in the biotope by photosynthesis. Because we eat plants, or animals that live off plants, the ¹⁴C content in the atmosphere is directly mirrored in the human body.

Radiocarbon dating has been used to study the incorporation of atmospheric ¹⁴C into DNA to determine the age of different human cell types, including adipocytes^{7–11}. Here, we compared the incorporation of ¹⁴C into adipocyte triglycerides with the dynamic changes in atmospheric ¹⁴C described earlier. Triglyceride age was determined by using a linear lipid replacement model in which the age distribution of lipids within an individual was exponentially distributed corresponding to a constant turnover rate (per year)¹². The associated mean age, termed lipid age, is the inverse of the turnover rate and reflects the irreversible removal of lipids from adipose stores (Supplementary Information 1 and Fig. 1 of Supplementary Information 1).

Earlier studies indicate that triglycerides in adipose tissue form two distinct pools with high or low turnover rates, respectively^{13,14}. Our data, obtained from individuals born before, during and after bomb testing, do not support the hypothesis of dual large lipid pools with different half-lives (Fig. 1b). ¹⁴C data were modelled according to one or more pools of lipids with different lipid removal rates (Supplementary Information 1). The existence of a very small pool of younger lipids cannot be excluded based on data modelling (Supplementary Information 1 and Fig. 2 of Supplementary Information 1). According to a two-pool model the influence on the turnover rate is proportional to the fraction of lipid in the small pool. Triglyceride exchange between adipocytes and other small storage pools can affect turnover estimates. The two-pools model shows, however, that the non-adipose pool can be neglected when it makes up less than 20% of the lipids (Supplementary Information 1, Fig. 3). Small pools with high turnover are more important for short-term (days or weeks) than long-term (years) triglyceride turnover.

Mean lipid age was 1.6 years (Fig. 1c), which is in the same range as in short-term turnover studies⁴. The distribution of lipid age was compared with that of adipocyte age reported previously in a comparable cohort⁹. The mean age of adipocytes was 9.5 years (Fig. 1d). This implies that triglycerides, on average, are replaced six times during the lifespan of the adipocyte, enabling a dynamic regulation of lipid storage and mobilization over time.

There is a large variation in adipocyte size within and between individuals (Supplementary Information 2, Supplementary Table 1)¹⁵. However, it is unlikely that the rate of triglyceride removal from adipocytes is important for these variations, as lipid age was not related to adipocyte size when set in relation to the body fat mass (Fig. 2a, b), nor was there a difference in lipid age between large and small adipocytes of the same adipose tissue sample (Fig. 2c, d). These data indicate that there is a continuous exchange of lipids between adipocytes within the adipose tissue that is not dependent on adipocyte size. Fatty acids produced by lipolysis in one adipocyte could, for example, be taken up

¹Department of Medicine, Karolinska University Hospital, SE-141 86 Stockholm, Sweden. ²Institut Camille Jordan, CNRS UMR 5208, University of Lyon, F-69622 Villeurbanne, France. ³Department of Physics and Astronomy, Ion Physics, Uppsala University, SE-751 20, Sweden. ⁴Faculty of Physics – Isotope Research, University of Vienna, Vienna, A-1090, Austria. ⁵Center for Accelerator Mass Spectrometry, Lawrence Livermore National Laboratory, 7000 East Ave., L-397, Livermore, California 94551 USA. ⁶Elsa Kröner-Fresenius Centre for Nutritional Medicine, Technische Universität München, D-85350 Weihenstephan, Germany. ⁷Oxford Centre for Diabetes, Endocrinology & Metabolism, Churchill Hospital, Oxford OX3 7LJ, UK. ⁸Department of Cell and Molecular Biology, Karolinska Institute, SE-171 77, Stockholm, Sweden.

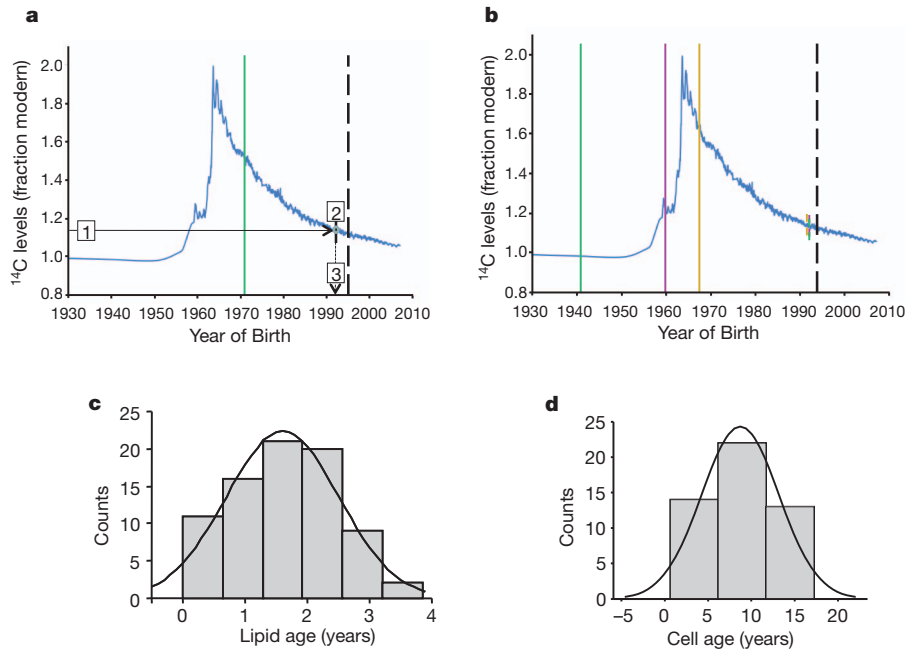


Figure 1 | Atmospheric ^{14}C over time and its use to determine lipid age and adipocyte age. **a**, Above ground nuclear bomb testing during the period of the cold war caused an increase in atmospheric levels of ^{14}C . These values decreased exponentially following implementation of a limited world-wide test ban treaty in 1963 (blue curve). Lipid age is determined by measuring ^{14}C levels in lipids (1) and plotting this value against the bomb curve (2) to determine the difference between the year corresponding to the atmospheric ^{14}C concentration (3) and the biopsy collection date (dashed line). Atmospheric ^{14}C levels are presented as $^{14}\text{C}/^{12}\text{C}$ ratios in units of fraction modern (for a definition of ‘modern’ see Supplementary Information 2). **b**, Lipid age and

turnover do not change as a function of person age. Lipid age is shown for three individuals born in 1940.2, 1959.9 and 1967.9. Lipid age was shown to be the same for all individuals, despite markedly different subject ages. Fat biopsies were collected from all individuals on the same date (dashed vertical line). The solid vertical lines indicate the date of birth. The small dashed lines show the ^{14}C lipid value for each individual. **c**, Distribution of values for lipid age in healthy non-obese or obese individuals from cohort 1 ($n = 78$). **d**, The distribution of values for human adipocyte age ($n = 27$). Adipocyte age data are obtained from a previous publication (see main text).

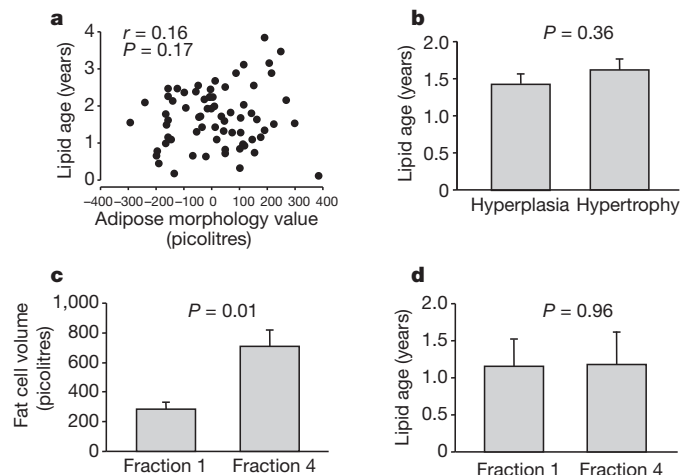


Figure 2 | Relationship between adipocyte size and lipid age. **a**, **b**, Influence of adipocyte cellularity on lipid age. Individuals were assigned a morphometric value, which is the difference between the measured adipocyte volume for the individual minus the average adipocyte volume for all subjects (see Supplementary Information 2). This analysis was carried out across the full range of body masses. Positive values indicate larger adipocytes than expected (fewer but larger adipocytes = hypertrophy). Negative values indicate smaller adipocytes than expected (many but smaller adipocytes = hyperplasia). **a**, Individual values compared by linear regression analysis ($n = 74$). **b**, Data (mean \pm standard error) with morphology as a dichotomous variable ($n = 36$ for hyperplasia and $n = 38$ for hypertrophy). An unpaired t -test was used. **c**, **d**, Isolated subcutaneous adipocytes were fractioned into small (fraction 1) or very large (fraction 4) samples ($n = 7$). Adipocyte volume (**c**) and adipose lipid age (**d**) were compared. Values are mean \pm standard error. A paired t -test was used. Data in **a** and **b** are from non-obese plus obese individuals in cohort 1 and data in **c** and **d** are from cohort 2.

by adjacent adipocytes and incorporated into their triglycerides. These processes would not be part of lipid removal as measured here.

Lipid age and total fat mass data were used to determine the net triglyceride storage in adipose tissue (kg year^{-1}) (see Supplementary Information 1). The net amount of lipid stored in adipose tissue each year is the sum of exogenous fat incorporation and endogenous synthesis, minus lipid removal. The removal rate represents the hydrolysis of triglycerides (lipolysis) followed by the irreversible removal of lipids by oxidation. A high lipid age therefore mirrors low removal rates. No relationship between lipid storage or removal and person age or gender was seen (Supplementary Information 2 and Fig. 1a–d of Supplementary Information 2).

Two clinical conditions where altered lipid metabolism is observed were investigated—obesity and familial combined hyperlipidaemia (FCHL); the latter is the most common hereditary lipid disorder (reviewed in ref. 16). It has an unknown aetiology and is a common hereditary cause of premature coronary heart disease. Adipocyte lipolysis is impaired in both conditions due to decreased cyclic AMP-dependent signalling, the major lipolytic pathway in adipocytes^{17–19}. Both conditions show a similar metabolic phenotype (mixed dyslipidaemia, elevated apolipoprotein B and insulin resistance)²⁰. These clinical characteristics are confirmed in our study cohort (Supplementary Information 2, Supplementary Table 1). FCHL individuals may present with a range of body fat levels; however, for our analyses only non-obese FCHL patients were selected so as to remove the confounding factor of obesity from the study.

In obese subjects, the rate of triglyceride storage (Fig. 3a) and mean lipid age (Fig. 3b) were markedly increased compared to non-obese individuals. Both lipid age ($r = 0.38$, $P = 0.0005$) and triglyceride storage ($r = 0.60$, $P < 0.0001$) correlated with body mass index (BMI) when non-obese and obese individual were pooled together. Similarly, in non-obese FCHL individuals lipid age was increased to values

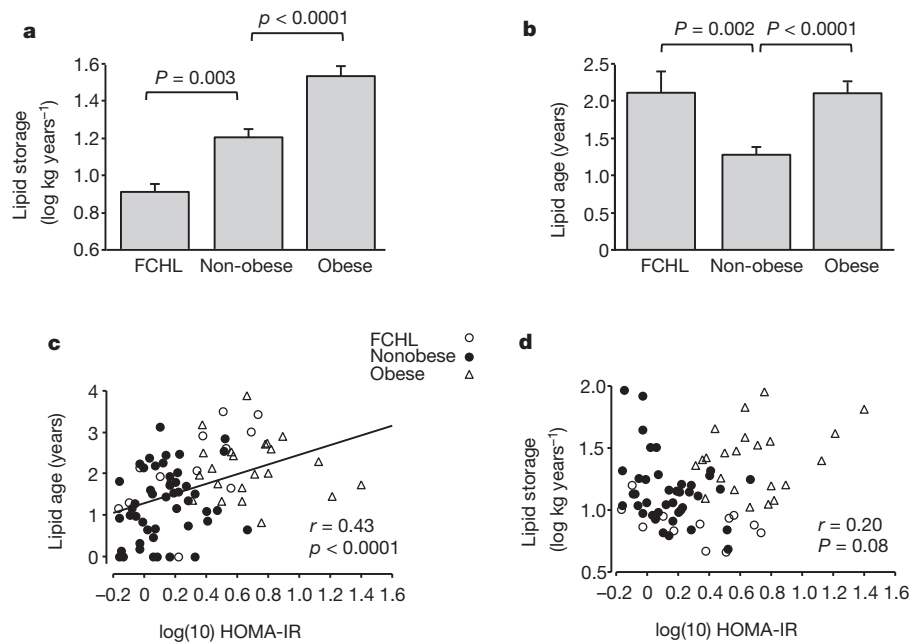


Figure 3 | Lipid turnover in subcutaneous fat. **a, b**, Lipid storage (**a**) and lipid age (**b**) were determined in 48 non-obese, 30 obese and 13 non-obese FCHL subjects. Error bars indicate standard error. Overall effect is $P < 0.0001$ by analysis of variance (ANOVA) in **a** and **b**. Results in graphs are from post-hoc test. Data are from cohort 1 (see Supplementary Information 2). A linear regression analysis was performed on all individuals from cohort 1 having

insulin resistance measures ($n = 82$). **c, d**, HOMA-IR was correlated with lipid age (**c**) and lipid storage (**d**). The relationship between lipid age and HOMA-IR remained significant when BMI, gender or group (non-obese, obese, FCHL) were included in the analysis (partial $r = 0.41$, $P = 0.006$ with BMI using multiple regression analysis and $F = 16.6$, $P = 0.0001$ and $F = 4.8$, $P = 0.03$ for gender or group, respectively, using analysis of covariance (ANCOVA)).

observed in obesity (Fig. 3b). In contrast to obesity, however, the rate of triglyceride storage was markedly decreased compared to non-obese individuals (Fig. 3a). Thus, adipocyte triglyceride turnover is not just a mere reflection of the fat mass. Our data indicate a model where a combination of high storage and low lipid removal rates, as in obesity, facilitates triglyceride accumulation within adipose tissue, thereby promoting the development and/or maintenance of excess body fat mass.

Conversely, a low rate of both triglyceride storage and removal, as in FCHL, leads to reduced triglyceride turnover and thereby a decreased ability of adipocytes to store and release fatty acids, despite a normal body fat mass. As discussed in detail elsewhere^{21,22}, low lipid turnover in adipose tissue may result in fatty acids being shunted to the liver, which drives the synthesis of apolipoprotein B and increases the circulating levels of triglycerides. Adipocyte triglyceride turnover may also be

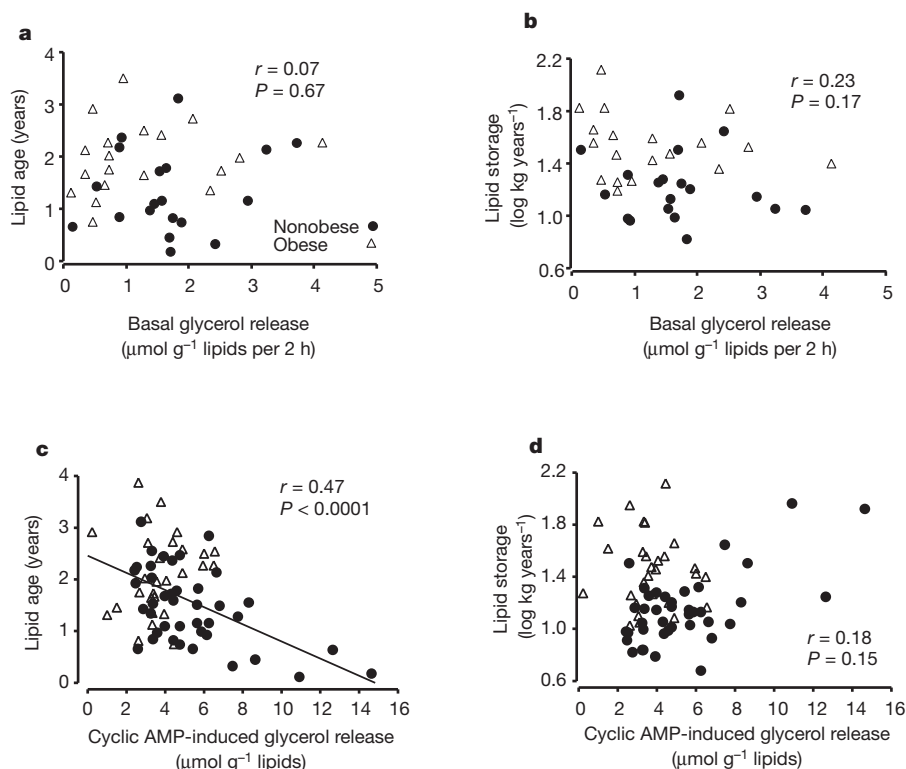


Figure 4 | Correlation between lipid turnover and adipocyte lipolysis. **a–d**, Lipid age and lipid removal rates were compared with basal rate of glycerol release (**a, b**) and with the rate of glycerol release induced with dibutyryl cyclic AMP (**c, d**), which is a phosphodiesterase-resistant and stable cyclic AMP analogue stimulating the protein kinase A complex. Linear regression analysis was used. Data are from non-obese ($n = 48$) and obese ($n = 28$) individuals from cohort 1. Data with dibutyryl cyclic AMP-induced lipolysis versus lipid age were also significant when analysed using BMI as a covariate in multiple regression analysis (partial $r = -0.40$; $P = 0.0006$).

involved in determining overall insulin effects. Insulin resistance (indirectly measured by the HOMA-IR index, see Supplementary Information 2) and lipid turnover were assessed in 82 individuals. Triglyceride age was strongly related to levels of insulin resistance (Fig. 3c), although there was no relationship between triglyceride storage and insulin resistance (Fig. 3d). There was no significant interaction between groups (lean, obese and non-obese FCHL) as determined by analysis of co-variance, indicating that the rate of triglyceride removal from adipocytes has an impact on whole-body insulin sensitivity independent of any underlying disorder. Multiple regression analysis showed that the relationship between HOMA-IR and lipid removal was not influenced by plasma triglycerides (partial $r = 0.35$; $P = 0.007$).

We also examined non-obese and obese individuals separately (Supplementary Information 2 and Figure 2a–d of Supplementary Information 2). Variations in BMI were significantly related to lipid age only among non-obese and to lipid storage only among obese individuals. HOMA-IR variations were significantly related to lipid storage when obese subjects were removed from the analysis (no relationship was found among obese subjects themselves). Thus, variations in triglyceride turnover may have a different impact on metabolic status in obese versus non-obese populations. Clearly, this assumption must be confirmed by investigations in much larger samples.

Because adipose tissue lipolysis is the first step in lipid removal, we investigated the ability of the cyclic AMP system to activate lipolysis *in vitro* in adipocytes isolated from lean and obese individuals and compared this with *in vivo* measurements of lipid storage and removal. Spontaneous (basal) lipolysis was not related to lipid turnover (Fig. 4a, b). However, the stimulated rate of lipolysis was positively correlated with triglyceride removal (inversely correlated with lipid age) but was not related to the rate of triglyceride uptake (lipid storage). This was irrespective of whether lipolysis was induced using a cyclic AMP analogue (Fig. 4c, d), by activating endogenous adenylate cyclase (using forskolin; Supplementary Information 2 and Fig. 3a, b of Supplementary Information 2) or by administration of a synthetic β -adrenoceptor-selective catecholamine (isoprenaline; Fig. 3c, d of Supplementary Information 2). These data indicate that lipolysis determines lipid turnover in adipocytes by regulating the rate of triglyceride removal. The impact of subsequent fatty acid oxidation could not be examined in this study; however, decreased lipid oxidation is frequently observed in obesity^{23,24}. As there are regional variations in lipolysis and all our studies were performed on one fat depot no attempts were made to extrapolate findings to the whole-body level.

We are in the midst of a global epidemic of obesity with negative health and socio-economic consequences. We propose adipose triglyceride turnover as a novel target for the prevention and treatment of excess body fat and possibly its consequences for insulin resistance. New insights into abnormal triglyceride turnover in FCHL patients may also suggest novel treatment strategies for this complex disease that targets adipocytes.

METHODS SUMMARY

Subjects. Subcutaneous adipose tissue was obtained from two patient cohorts. Patient selection and collection of clinical data are described in Supplementary Information 2.

Preparation of lipids. Triglycerides were extracted from pieces of adipose tissue or isolated adipocytes. Details of lipid extraction and adipocyte isolation are given in Supplementary Information 2. Extracted lipids were subjected to accelerator mass spectrometry analysis, as described in Supplementary Information 2.

Data analysis. Calculations between lipid turnover and clinical or adipocyte phenotypes are described in Supplementary Information 2. Calculations of lipid age and net lipid uptake by adipose tissue are described in Supplementary Information 1. Conventional statistical methods were used to summarize and compare data.

Received 24 June; accepted 5 August 2011.

Published online 25 September 2011.

1. Klein, R. A., Halliday, D. & Pittet, P. G. The use of 13-methyltetradecanoic acid as an indicator of adipose tissue turnover. *Lipids* **15**, 572–579 (1980).

2. Mårin, P., Oden, B. & Björntorp, P. Assimilation and mobilization of triglycerides in subcutaneous abdominal and femoral adipose tissue *in vivo* in men: effects of androgens. *J. Clin. Endocrinol. Metab.* **80**, 239–243 (1995).
3. Mårin, P., Rebuffe-Scrive, M. & Björntorp, P. Uptake of triglyceride fatty acids in adipose tissue *in vivo* in man. *Eur. J. Clin. Invest.* **20**, 158–165 (1990).
4. Strawford, A., Antelo, F., Christiansen, M. & Hellerstein, M. K. Adipose tissue triglyceride turnover, *de novo* lipogenesis, and cell proliferation in humans measured with $2\text{H}_2\text{O}$. *Am. J. Physiol.* **286**, E577–E588 (2004).
5. Nydal, R. & Lovseth, K. Distribution of radiocarbon from nuclear tests. *Nature* **206**, 1029–1031 (1965).
6. Levin, I. & Kromer, B. The tropospheric $^{14}\text{CO}_2$ level in mid latitudes of the northern hemisphere (1959–2003). *Radiocarbon* **46**, 1261–1272 (2004).
7. Spalding, K. L., Bhardwaj, R. D., Buchholz, B. A., Druid, H. & Frisen, J. Retrospective birth dating of cells in humans. *Cell* **122**, 133–143 (2005).
8. Perl, S. *et al.* Significant human β -cell turnover is limited to the first three decades of life as determined by *in vivo* thymidine analog incorporation and radiocarbon dating. *J. Clin. Endocrinol. Metab.* **95**, E234–E239 (2010).
9. Spalding, K. L. *et al.* Dynamics of fat cell turnover in humans. *Nature* **453**, 783–787 (2008).
10. Bergmann, O. *et al.* Evidence for cardiomyocyte renewal in humans. *Science* **324**, 98–102 (2009).
11. Bhardwaj, R. D. *et al.* Neocortical neurogenesis in humans is restricted to development. *Proc. Natl Acad. Sci. USA* **103**, 12564–12568 (2006).
12. Bernard, S., Frisen, J. & Spalding, K. L. A mathematical model for the interpretation of nuclear bomb test derived ^{14}C incorporation in biological systems. *Nucl. Instrum. Meth. B* **268**, 1295–1298 (2010).
13. Ekstedt, B. & Olivecrona, T. Uptake and release of fatty acids by rat adipose tissue: last in to first out? *Lipids* **5**, 858–860 (1970).
14. Kerpel, S., Shafir, E. & Shapiro, B. Mechanism of fatty acid assimilation in adipose tissue. *Biochim. Biophys. Acta* **46**, 495–504 (1961).
15. Björntorp, P. Effects of age, sex, and clinical conditions on adipose tissue cellularity in man. *Metabolism* **23**, 1091–1102 (1974).
16. Grundy, S. M., Chait, A. & Brunzell, J. D. Familial combined hyperlipidemia workshop. *Arterioscler. Thromb. Vasc. Biol.* **7**, 203–207 (1987).
17. Langin, D. *et al.* Adipocyte lipases and defect of lipolysis in human obesity. *Diabetes* **54**, 3190–3197 (2005).
18. Reynisdottir, S., Eriksson, M., Angelin, B. & Arner, P. Impaired activation of adipocyte lipolysis in familial combined hyperlipidemia. *J. Clin. Invest.* **95**, 2161–2169 (1995).
19. van der Kallen, C. J. *et al.* Evidence of insulin resistant lipid metabolism in adipose tissue in familial combined hyperlipidemia, but not type 2 diabetes mellitus. *Atherosclerosis* **164**, 337–346 (2002).
20. Ayyobi, A. F. & Brunzell, J. D. Lipoprotein distribution in the metabolic syndrome, type 2 diabetes mellitus, and familial combined hyperlipidemia. *Am. J. Cardiol.* **92**, 27–33 (2003).
21. Arner, P. Is familial combined hyperlipidaemia a genetic disorder of adipose tissue? *Curr. Opin. Lipidol.* **8**, 89–94 (1997).
22. de Graaf, J., Veerkamp, M. J. & Stalenhoef, A. F. Metabolic pathogenesis of familial combined hyperlipidaemia with emphasis on insulin resistance, adipose tissue metabolism and free fatty acids. *J. R. Soc. Med.* **95** (suppl. 42), 46–53 (2002).
23. Blaak, E. E. *et al.* Fat oxidation before and after a high fat load in the obese insulin-resistant state. *J. Clin. Endocrinol. Metab.* **91**, 1462–1469 (2006).
24. Houmard, J. A. Intramuscular lipid oxidation and obesity. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **294**, R1111–R1116 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This study was supported by the Swedish Research Council, Swedish Foundation for Strategic Research, Swedish Heart and Lung Foundation, Novo Nordic Foundation, Swedish Diabetes Foundation, Strategic Research Program in Diabetes at the Karolinska Institutet, Swedish Cancer Society, Uppsala BIO, Sweden, NIH/NCRR Grant RR13461, ERC grant 261258-HUFATREG and by the projects 'Hepatic and adipose tissue and functions in the metabolic syndrome' (HEPADIP, <http://www.hepadip.org/>) and 'Adipokines as drug targets to combat adverse effects of excess adipose tissue' (ADAPT, <http://www.adapt-eu.net/>), which were supported by the European Commission as an Integrated Project under the 6th and the 7th Framework Programmes (contract LSHM-CT-2005-018734 and contract HEALTH-F2-2008-201100). This work was performed in part under the auspices of the US Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. The authors would like to acknowledge E. Sjölin, K. Wåhlén, B.-M. Leijonhufvud, K. Hertel and Y. Widlund for technical assistance. We would like to thank F. Barnabé-Heider and J. Frisen for useful comments on the manuscript.

Author Contributions K.L.S. and P.A. designed the study and wrote the manuscript together with K.N.F. and S.B. M.R. co-ordinated writing and data assembly. S.B. and E.A. were responsible for the modelling. K.L.S. performed sample preparation. M.S., G.P., B.A.B., P.S. and J.L. performed ^{14}C accelerator mass spectrometry measurements. P.A., M.E., T.S. and H.H. collected clinical material.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to P.A. (peter.arnar@ki.se) or K.L.S. (kirsty.spalding@ki.se).

An olfactory receptor for food-derived odours promotes male courtship in *Drosophila*

Yael Grosjean^{1,2}, Raphael Rytz¹, Jean-Pierre Farine², Liliane Abuin¹, Jérôme Cortot², Gregory S. X. E. Jefferis³ & Richard Benton¹

Many animals attract mating partners through the release of volatile sex pheromones, which can convey information on the species, gender and receptivity of the sender to induce innate courtship and mating behaviours by the receiver¹. Male *Drosophila melanogaster* fruitflies display stereotyped reproductive behaviours towards females, and these behaviours are controlled by the neural circuitry expressing male-specific isoforms of the transcription factor Fruitless (FRU^M)^{2–5}. However, the volatile pheromone ligands, receptors and olfactory sensory neurons (OSNs) that promote male courtship have not been identified in this important model organism. Here we describe a novel courtship function of Ionotropic receptor 84a (IR84a), which is a member of the chemosensory ionotropic glutamate receptor family⁶, in a previously uncharacterized population of FRU^M -positive OSNs. IR84a-expressing neurons are activated not by fly-derived chemicals but by the aromatic odours phenylacetic acid and phenylacetaldehyde, which are widely found in fruit and other plant tissues⁷ that serve as food sources and oviposition sites for drosophilid flies⁸. Mutation of *Ir84a* abolishes both odour-evoked and spontaneous electrophysiological activity in these neurons and markedly reduces male courtship behaviour. Conversely, male courtship is increased—in an IR84a-dependent manner—in the presence of phenylacetic acid but not in the presence of another fruit odour that does not activate IR84a. Interneurons downstream of IR84a-expressing OSNs innervate a pheromone-processing centre in the brain. Whereas IR84a orthologues and phenylacetic-acid-responsive neurons are present in diverse drosophilid species, IR84a is absent from insects that rely on long-range sex pheromones. Our results suggest a model in which IR84a couples food presence to the activation of the fru^M courtship circuitry in fruitflies. These findings reveal an unusual but effective evolutionary solution to coordinate feeding and oviposition site selection with reproductive behaviours through a specific sensory pathway.

While mapping the projections of Ionotropic receptor (IR)-expressing OSNs to the primary olfactory centre⁹, the antennal lobe, we observed that an *Ir84a* reporter was labelling neurons innervating the VL2a glomerulus. VL2a is one of only three glomeruli that are larger in males and whose OSN inputs and projection neuron outputs express male-specific isoforms of the behavioural sex determination gene *fruitless* (fru^M)^{4,5}. fru^M -expressing OSNs have been implicated in promoting male sexual behaviours, because inhibition of synaptic transmission in all of these neurons simultaneously reduces male courtship of females⁵. We confirmed the expression of *Ir84a* in fru^M -expressing neurons by visualizing the co-expression of an *Ir84a* reporter, as well as endogenous *Ir84a* transcripts, with a fru^M reporter^{4,5} (Fig. 1a and Supplementary Fig. 1a). No sexual dimorphism was observed either in the number of *Ir84a*-expressing cells (males, 30 ± 2 neurons; females, 30 ± 3 neurons; $n = 4$ antennae each) or in their targeting to VL2a, indicating that FRU^M does not have an essential role in the development of these neurons, similar to other fru^M -expressing OSNs⁵.

We generated a *GAL4* knock-in null allele, *Ir84a^{GAL4}* (Fig. 1b, c). *Ir84a^{GAL4/+}* heterozygotes expressed a *GAL4*-responsive, membrane-targeted, green fluorescent protein (GFP) transgene (*UAS-mCD8:GFP*) exclusively in *Ir84a*-expressing OSNs (Fig. 1d, *Ir84a^{+/-}*, and Supplementary Fig. 1b). In *Ir84a^{GAL4}* homozygotes, the endogenous expression of *Ir84a* was lost, but the distribution and dendritic projections of these neurons, as revealed by mCD8:GFP, was unaffected (Fig. 1d, *Ir84a^{-/-}*). The axons of *Ir84a*-expressing neurons in heterozygous and homozygous *Ir84a^{GAL4}* flies projected only to VL2a (Fig. 1e). *Ir84a* is therefore dispensable for the specification and wiring of the neurons in which it is expressed. An amino-terminal enhanced GFP (EGFP)-tagged version of this receptor¹⁰ localized to the cell bodies and the ciliated dendritic endings of these neurons but not to their axon termini (Fig. 1f, g), consistent with an exclusive role for IR84a as an olfactory receptor in the fru^M circuitry.

We tested the responses of IR84a-expressing neurons to chemicals produced by male or virgin female flies¹¹, both by delivering headspaces of flies from a distance (simulating the action of volatile pheromones) and by presenting extracts from fly cuticles at close range (mimicking exposure to non-volatile hydrocarbons, such as contact pheromones¹²). These stimuli produced no or extremely small responses, as detected by extracellular recordings in ac4 sensilla (Supplementary Fig. 2a), which belong to the class of olfactory hair that houses IR84a-expressing neurons, as well as OSNs that express IR75d, or IR76a and IR76b⁶. These observations suggest that IR84a is not tuned to fly-derived pheromones. We therefore tested 163 structurally diverse odours. Only three of these gave responses of >50 spikes s^{-1} above basal activity: phenylacetaldehyde (as identified previously¹³), phenylacetic acid and phenylethylamine (Fig. 2a and Supplementary Fig. 2b). Dose response curves that revealed sensitivity to these ligands are similar in both sexes (Fig. 2b and data not shown).

In *Ir84a^{GAL4}* homozygous mutants, the responses to phenylacetic acid and phenylacetaldehyde were completely abolished (Fig. 2c, d, *Ir84a^{-/-}*). Re-introduction of *Ir84a* function in these neurons, by using *UAS-Ir84a* or *UAS-EGFP:Ir84a* cDNA transgenes, rescued these phenotypes (Fig. 2c, d, rescue), indicating a cell-autonomous function for IR84a in mediating these odour responses. By contrast, responses to phenylethylamine were unaffected (Fig. 2c, d), corroborating the evidence that this chemical is detected by the neurons that express both IR76a and IR76b^{6,10}. Consistent with these loss-of-function data, mis-expression of IR84a in Odorant receptor 35a (OR35a)-expressing neurons was sufficient to confer responsiveness to phenylacetic acid and phenylacetaldehyde (Supplementary Fig. 2c). The basal activity in *Ir84a* mutant ac4 sensilla was also lower than that in the ac4 sensilla of wild-type and rescue genotypes (Fig. 2e, f), indicating that IR84a has a role in promoting spontaneous firing.

Phenylacetic acid and phenylacetaldehyde are aromatic compounds found in a diverse range of fruit and other plant tissues⁷, as well as in their fermentation products¹⁴, and they are used in human perfumes for their floral, honey-like, sweet smell. We confirmed the presence of these chemicals in two host fruit for drosophilid flies⁸, overripe bananas and

¹Center for Integrative Genomics, Faculty of Biology and Medicine, University of Lausanne, CH-1015 Lausanne, Switzerland. ²Centre des Sciences du Goût et de l'Alimentation, UMR-6265 CNRS, UMR-1324 INRA, Université de Bourgogne, 6 Boulevard Gabriel, 21000 Dijon, France. ³Division of Neurobiology, MRC Laboratory of Molecular Biology, Cambridge CB2 0QH, UK.

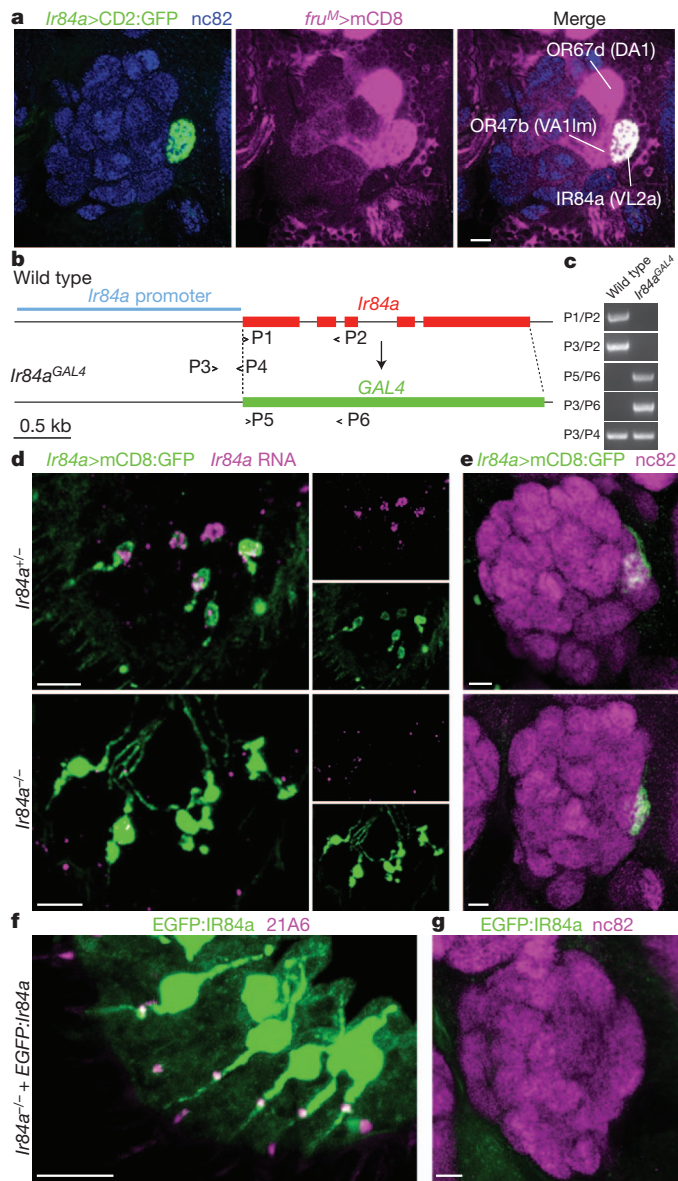


Figure 1 | Gene targeting of *Ir84a*, a candidate olfactory receptor in the *fru^M* circuit. **a**, Immunofluorescence using anti-GFP (green), anti-CD8 (magenta) and nc82 (blue; neuropil marker) antibodies on an antennal lobe of an animal in which an *Ir84a* promoter-*lexA* transgene drives (>) expression of a LexA-responsive CD2:GFP reporter transgene and in which the *fru^M* GAL4 insertion allele drives expression of a GAL4-responsive CD8 reporter transgene (genotype, *Ir84a-lexA/UAS-mCD8;lexAop-rCD2:GFP/fru^{GAL4}*). **b**, Schematic of the *Ir84a* locus, illustrating the organization of wild-type *Ir84a* and null mutant *Ir84a^{GAL4}* alleles. The span of the cloned *Ir84a* promoter is also indicated. P denotes PCR primer position. kb, kilobases. **c**, PCR confirmation of the *Ir84a^{GAL4}* allele using the pairs of primers shown in **b**. **d**, Combined *Ir84a* RNA *in situ* hybridization (magenta) and anti-GFP immunofluorescence (green) on antennae from *Ir84a* mutant flies: top, heterozygous mutants (*UAS-mCD8:GFP/+;Ir84a^{GAL4}/+*); and, bottom, homozygous mutants (*UAS-mCD8:GFP/+;Ir84a^{GAL4}/Ir84a^{GAL4}*). Separate channels are shown on the right. **e**, Immunofluorescence using anti-GFP (green) and nc82 (magenta) antibodies on antennal lobes from *Ir84a* mutant flies: top, control heterozygous mutants (*UAS-mCD8:GFP;Ir84a^{GAL4}/+*); and, bottom, homozygous mutants (*UAS-mCD8:GFP;Ir84a^{GAL4}/Ir84a^{GAL4}*). **f**, Immunofluorescence using anti-GFP antibodies (green) and antibodies specific for the sensory cilium base marker 21A6 (magenta) on an antenna expressing EGFP:Ir84a (*UAS-EGFP:Ir84a/+;Ir84a^{GAL4}/Ir84a^{GAL4}*). **g**, Immunofluorescence using anti-GFP (green) and nc82 (magenta) antibodies on an antennal lobe of a fly expressing EGFP:Ir84a (*UAS-EGFP:Ir84a;Ir84a^{GAL4}/Ir84a^{GAL4}*). **a**, **d**–**g**, Scale bar, 10 μ m.

the prickly-pear cactus *Opuntia ficus-indica*, as well as in laboratory *Drosophila* medium, by using gas chromatography–mass spectrometry analysis (Supplementary Table 1). The ubiquity of phenylacetic acid in vegetal tissues may be linked with its activity as a growth-regulating auxin⁷ and/or its production by plant-associated microorganisms¹⁵. Small, but reproducible, quantities of phenylacetic acid and phenylacetaldehyde were also detected in whole-body cuticular extracts of male and virgin female *D. melanogaster* (Supplementary Table 1). The similarity in the relative amounts of these chemicals in laboratory medium and fruitfly extracts (Supplementary Table 1) suggested that these chemicals are transferred from food to flies during their culture. ‘Clean’ cuticular extracts from animals grown on a minimal medium containing only sucrose and agarose consistently contain no detectable phenylacetaldehyde or phenylacetic acid (Supplementary Table 1).

The expression of IR84a in *fru^M*-expressing neurons implicates this receptor in the regulation of male courtship^{4,5}. Indeed, in single-pair courtship assays, we observed that *Ir84a^{GAL4}* mutant males court wild-type females significantly less than do wild-type males (Fig. 3a and Supplementary Fig. 3a, b). This phenotype was observed using both decapitated virgin females (which do not produce feedback signals) (Fig. 3a) and in more natural conditions, with intact females together with food (Supplementary Fig. 3b). Most individual components of the courtship ritual were affected in *Ir84a* mutant flies (Supplementary Fig. 3a). These defects were rescued with a *UAS-Ir84a* transgene, confirming that they result from the absence of IR84a in OSNs (Fig. 3a and Supplementary Fig. 3a, b). The observed reduction in male heterosexual courtship index (~50%) is highly comparable to the phenotype of flies in which all FRU^M-positive OSNs are silenced⁵, suggesting that IR84a-expressing neurons are the major olfactory *fru^M* channel contributing to this behaviour. Residual courtship is presumably stimulated by other sensory modalities, such as taste^{16,17}. Male wild-type *D. melanogaster* also show a low level of courtship towards other males, and we found that this homosexual courtship was also markedly reduced in *Ir84a^{GAL4}* mutants (Fig. 3b). By contrast, *Ir84a^{GAL4}* mutant females did not show overt defects in reproductive behaviours, including copulation latency, success or duration (Fig. 3c).

In innate olfactory preference assays, *Ir84a^{GAL4}* mutant flies still show robust avoidance of acetic acid, indicating that they do not have a general impairment in sensory detection (Supplementary Fig. 3c). By contrast, we did not observe any obvious responses of flies to phenylacetic acid (Supplementary Fig. 3d), suggesting that this food-derived odour is not a volatile stimulus that attracts flies but is a salient cue at close range. Notably, phenylacetic acid has a low vapour pressure (~0.005 mm Hg at 25 °C) compared with other fruit volatiles (for example, ethyl butyrate, which has a vapour pressure of 15 mm Hg at 25 °C). The observation that courtship is reduced in *Ir84a^{GAL4}* mutants in assays in which only small amounts of phenylacetic acid are present on fly cuticles (Supplementary Table 1) raises the possibility that spontaneous activity of these neurons also contributes to establishing a basal courtship level, which is abolished in the absence of IR84a (Fig. 2e, f).

To test whether IR84a ligands are sufficient to promote courtship, we adapted our assay by using killed female objects (which males court at only low levels) and by replacing the base of the chamber with gauze, beneath which we placed a filter paper treated with odour or solvent. Perfuming with phenylacetic acid nearly doubled the courtship index of wild-type flies compared with a solvent control (Fig. 3d). This effect was abolished in *Ir84a^{GAL4}* mutants and could be restored, albeit not fully, by introducing a *UAS-Ir84a* transgene. By contrast, ethyl butyrate, which does not activate IR84a (Supplementary Fig. 2b), did not increase courtship (Fig. 3d). We also perfumed the courtship chamber with *Drosophila* food—which contains phenylacetic acid (Supplementary Table 1)—and observed that this complex olfactory stimulus induced IR84a-dependent increases in male courtship behaviour (Fig. 3e).

The other *fru^M*-expressing OSN populations express either OR67d, which is a receptor for the antiaphrodisiac male pheromone *cis*-

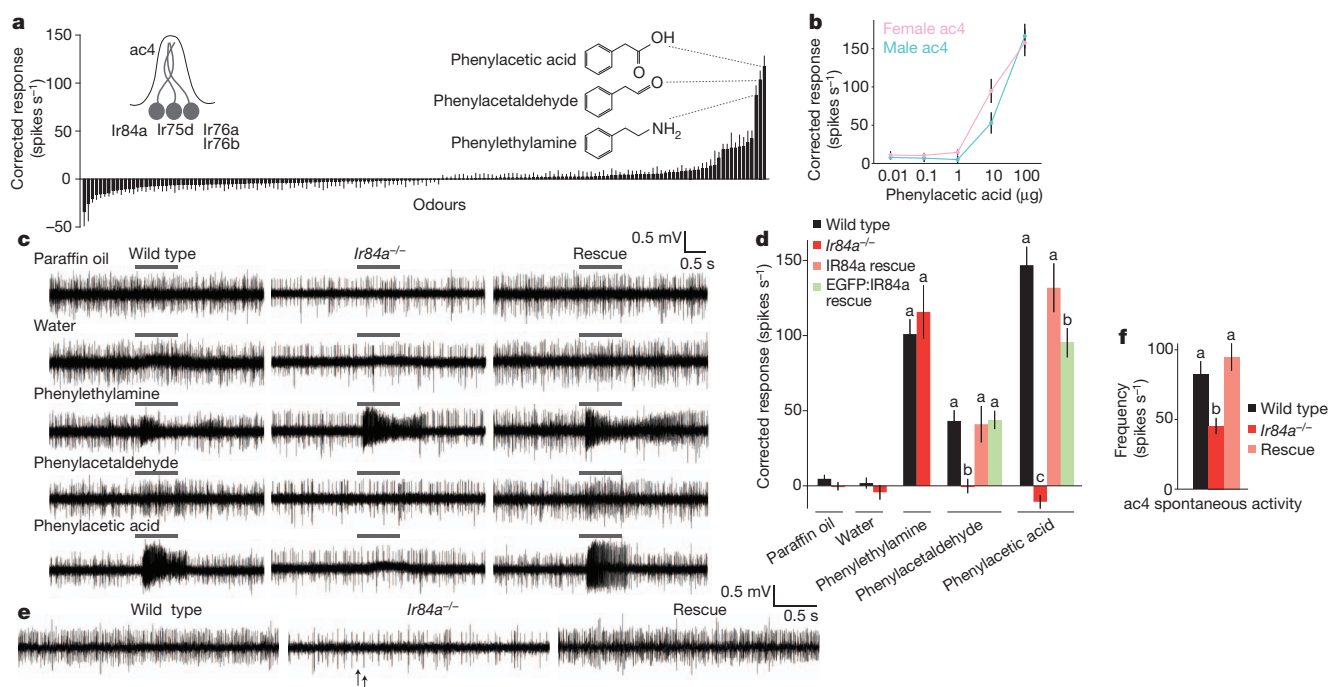


Figure 2 | Essential requirement for IR84a for spontaneous and odour-evoked neuronal responses. **a**, An odour ligand screen in ac4 sensilla. The mean evoked responses of ac4 sensilla neurons to a panel of 163 odours (representing the summed activities of the IR84a-expressing, IR75d-expressing, and IR76a- plus IR76b-expressing neurons). Data are presented as mean ± s.e.m.; *n* ≥ 4, mixed genders. An enlarged version of this histogram with all of the odour names indicated is provided in Supplementary Fig. 2b. The structures of the best three agonists are shown. **b**, The dose response curve to phenylacetic acid of male and female ac4 sensilla. Data are presented as mean ± s.e.m.; *n* = 12 for each gender. The doses reflect the total mass of phenylacetic acid (diluted in solvent) in the odour syringe. The responses of the male and female sensilla are not significantly different at any concentration (analysis of variance (ANOVA), *P* > 0.0571). A higher concentration of phenylacetic acid could not be prepared because of its limited solubility. **c**, Representative traces of recordings of ac4 sensilla stimulated with the indicated odour stimuli (10 μl phenylacetic acid (10 μg μl⁻¹) or 10 μl (1% (v/v)) other odours) in wild-type (*w¹¹¹⁸*), *Ir84a*^{GAL4}/*Ir84a*^{GAL4} mutant (*Ir84a^{GAL4}/Ir84a^{GAL4}*) and *Ir84a* rescue (*UAS-Ir84a/+; Ir84a^{GAL4}/Ir84a^{GAL4}*) flies. Paraffin oil and water

are solvent controls. The grey bars above the traces mark the stimulus time (1 s). **d**, Quantification of mean responses to the indicated odours for the experiment shown in **c**. Data are presented as mean ± s.e.m.; *n* ≥ 11, mixed genders. The restoration of phenylacetaldehyde and phenylacetic acid responses by the expression of EGFP:IR84a (*UAS-EGFP:Ir84a/+; Ir84a^{GAL4}/Ir84a^{GAL4}*) is also shown (green). For each odour stimulus, bars labelled with different letters are significantly different from each other (phenylethylamine ANOVA, *P* = 0.4861; phenylacetaldehyde ANOVA, *P* = 0.0003; phenylacetic acid ANOVA, *P* < 0.0001). For phenylacetic acid, EGFP:IR84a rescue is partial. **e**, Representative traces of spontaneous activity in ac4 sensilla in wild-type, *Ir84a*^{GAL4}/*Ir84a*^{GAL4} mutant (*Ir84a^{GAL4}/Ir84a^{GAL4}*) and *Ir84a* rescue (*UAS-Ir84a/+; Ir84a^{GAL4}/Ir84a^{GAL4}*) flies. The arrows mark two distinct spike amplitudes discernible in the sparser spike trains in *Ir84a* mutant sensilla; these are likely to correspond to the IR75d- and the IR76a- plus IR76b-expressing neurons. **f**, Quantification of the mean spontaneous activity in ac4 sensilla in the experiment shown in **e**. Data are presented as mean ± s.e.m.; *n* ≥ 12, mixed genders. Bars labelled with different letters are significantly different (ANOVA, *P* = 0.0005).

vaccenyl acetate¹⁸, or OR47b, which is activated by unidentified fly-derived odours from both sexes¹¹ and may participate in mate localization^{19,20}. We examined how IR84a sensory information is integrated with these pheromonal pathways, by visualizing the axons of projection neurons innervating the VL2a (IR84a), VA1lm (OR47b) and DA1 (OR67d) glomeruli, which carry sensory information to the mushroom body and lateral horn²¹ (Fig. 4a). We registered images of single-labelled projection neurons of different glomerular classes onto a common reference brain (see Methods). DA1 and VA1lm excitatory projection neurons target an anterior-ventral pheromone-processing region of the lateral horn²², which is segregated from projection neurons that are responsive to general food odours²². Importantly, we found that VL2a projection neurons—and no other IR-expressing projection neuron class⁹—are highly interdigitated with pheromone pathways and not food pathways (Fig. 4a, b). Indeed, VL2a projection neuron axon terminals overlap more strongly with VA1lm projection neurons than any of the other 44 projection neuron classes (see Methods and Supplementary Fig. 4b, c), consistent with projection neurons of both of these classes transmitting courtship-promoting sensory signals^{19,20}. The VL2a, DA1 and VA1lm inhibitory projection neurons were observed to overlap to a similar extent (Supplementary Fig. 4a–c). The anatomical convergence of combinations of excitatory and inhibitory inputs from VL2a, VA1lm and DA1 projection neurons

may allow the integration of olfactory signals by *fru^M*-expressing third-order neurons^{23,24} to control male courtship behaviour.

Many olfactory IRs are conserved in insects²⁵ and may detect odours that are important for all species. By contrast, although IR84a orthologues are present in ecologically diverse drosophilids, they are absent from other Diptera and more divergent insects (Supplementary Fig. 5a). In the cactophilic species *Drosophila mojavensis*, we identified coeloconic sensilla with neurons that are responsive to phenylacetic acid and phenylacetaldehyde on their anterior antennal surface (similar to ac4 sensilla in *D. melanogaster*) (Supplementary Fig. 5b). Thus, IR84a may have a conserved, drosophilid-specific function.

Despite the widely held assumption of the existence of volatile chemicals that promote courtship in *Drosophila*^{4,5}, behavioural evidence for long-range pheromones is inconclusive^{26,27}, and no female-specific volatile compound that activates male OSNs has been identified¹¹. Our characterization of IR84a identifies an olfactory receptor that is expressed in *fru^M*-positive neurons and is required to promote male courtship. Surprisingly, this receptor is not activated by fly-derived odours but rather by aromatic compounds that are present in the vegetal substrates in which fruitflies feed, breed and oviposit⁸. Thus, the IR84a pathway may promote male courtship in the presence of food, complementing the functions of pheromone receptors in regulating mate choice (Supplementary Fig. 5c). This model can account for

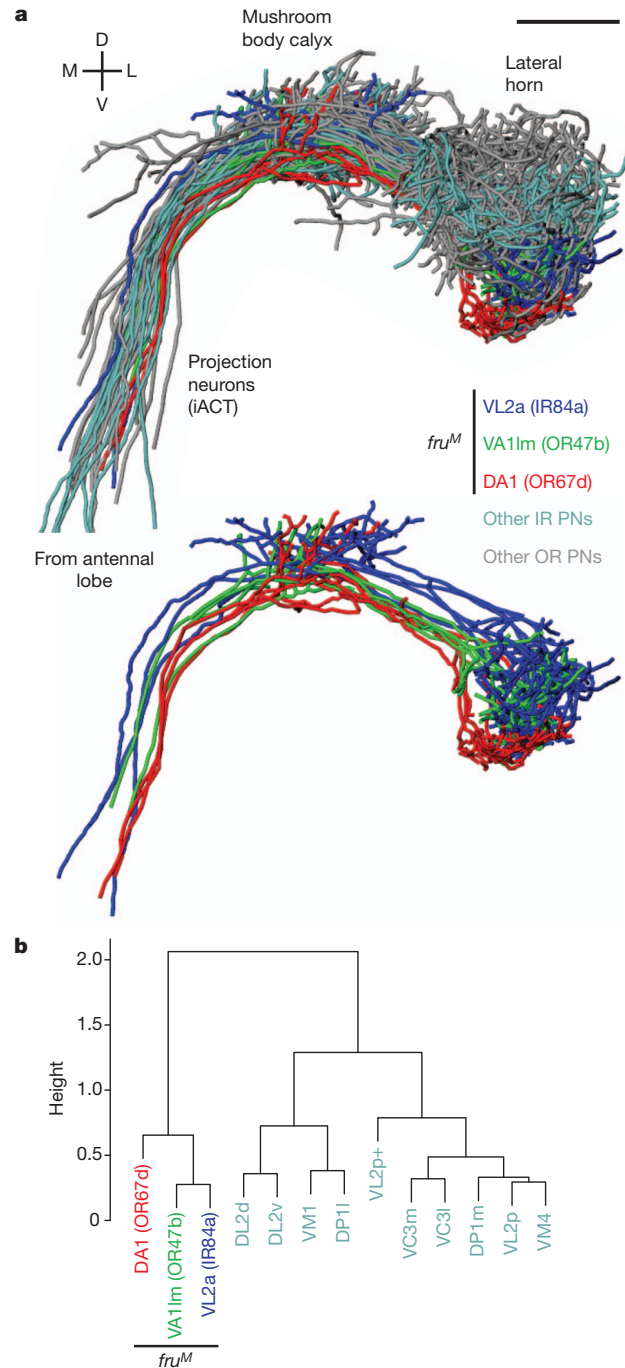
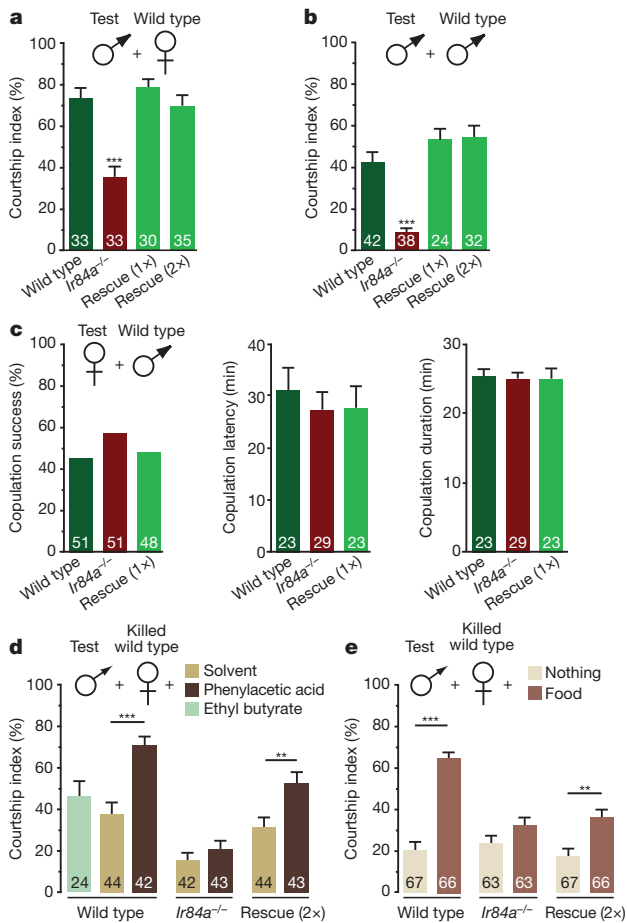


Figure 4 | Anatomical integration of VL2a ($Ir84a$ -expressing) projection neurons in the pheromone processing centre. **a**, Top, three-dimensional rendering of registered axonal projections of projection neurons (PNs) receiving input from the VL2a glomerulus ($Ir84a$ -expressing) (dark blue), the VA1Im glomerulus (OR47b-expressing) (green) and the DA1 glomerulus (OR67d-expressing) (red) ($n = 4$ for each class). Other classes of projection neuron are shown in cyan (nine IRs^9) and grey (27 ORs^{22}) ($n = 1$ for each class). These projection neurons are cholinergic and project to both the mushroom body and the lateral horn through the inner antennocerebral tract (iACT); a small minority of other projection neurons are GABAergic (γ -aminobutyric acid-containing) and project exclusively to the lateral horn through the middle ACT (Supplementary Fig. 4a). Bottom, when the other projection neurons are removed, the extensive co-mingling of VL2a and VA1Im projection neurons is evident. Scale bar, 25 μm . D, dorsal; L, lateral; M, medial; V, ventral. **b**, Cluster analysis of normalized axon overlap scores (y axis) using Ward’s method (see Methods) demonstrates that VL2a projection neurons cluster with projection neurons from other fru^M -expressing glomeruli rather than with other IR -expressing projection neurons. Colours are as in **a**.

the widespread observations that *D. melanogaster* and other drosophilids mate predominantly on their food substrates^{8,28}. Whereas many insects and other animal classes use long-range sex pheromones to attract potential mates¹, the evolution of IR84a in fruitflies has provided an alternative (although not necessarily exclusive) olfactory mechanism to unite males with females by integrating food-sensing neurons with the circuitry controlling sexual behaviour. Whether other animals have dedicated sensory pathways for environmental 'aphrodisiacs' remains an open question.

METHODS SUMMARY

Genetic and molecular biological manipulations and histological and electrophysiological analyses were performed essentially as described previously^{6,9,10}. Responses to *Drosophila*-derived stimuli were measured by adapting previous protocols¹¹. *Drosophila* extracts and food or fruit extracts were made using CH₂Cl₂ solvent and were analysed by gas chromatography–mass spectrometry on a GC-MS-QP2010 apparatus (Shimadzu). Courtship tests were performed at 25 °C in round observation chambers (1.2 cm in diameter and 0.4 cm deep) under far-red light and were video recorded and annotated manually by observers who were blinded to the genotypes. For perfumed behavioural assays, the courtship chamber was placed on top of an identical chamber, with the chambers separated by muslin gauze to avoid contact between the flies in the top chamber and the chemical stimuli placed in the bottom chamber. Olfactory responses were measured using a Y-maze assay. Projection neurons from several sources (see Methods) were traced, where necessary, using Amira software (<http://www.amira.com>) and registered, by adapting previous approaches²². The degree of overlap between the axon terminals of different projection neurons was measured using the three-dimensional convex hull formed by each set of terminals. Cluster analysis of these data was performed using Ward's method.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 4 June 2010; accepted 3 August 2011.

Published online 28 September 2011.

- Wyatt, T. D. *Pheromones and Animal Behaviour: Communication by Smell and Taste* (Oxford Univ. Press, 2003).
- Dickson, B. J. Wired for sex: the neurobiology of *Drosophila* mating decisions. *Science* **322**, 904–909 (2008).
- Demir, E. & Dickson, B. J. *fruitless* splicing specifies male courtship behavior in *Drosophila*. *Cell* **121**, 785–794 (2005).
- Manoli, D. S. *et al.* Male-specific *fruitless* specifies the neural substrates of *Drosophila* courtship behaviour. *Nature* **436**, 395–400 (2005).
- Stockinger, P., Kvitsiani, D., Rotkopf, S., Tirian, L. & Dickson, B. J. Neural circuitry that governs *Drosophila* male courtship behavior. *Cell* **121**, 795–807 (2005).
- Benton, R., Vannice, K. S., Gomez-Diaz, C. & Vosshall, L. B. Variant ionotropic glutamate receptors as chemosensory receptors in *Drosophila*. *Cell* **136**, 149–162 (2009).
- Wightman, F. & Lighty, D. L. Identification of phenylacetic acid as a natural auxin in the shoots of higher plants. *Physiol. Plant.* **55**, 17–24 (1982).
- Markow, T. A. & O'Grady, P. Reproductive ecology of *Drosophila*. *Funct. Ecol.* **22**, 747–759 (2008).
- Silbering, A. F. *et al.* Complementary function and integrated wiring of the evolutionarily distinct *Drosophila* olfactory subsystems. *J. Neurosci.* doi:10.1523/JNEUROSCI.2360-11.2011 (21 September 2011).
- Abuin, L. *et al.* Functional architecture of olfactory ionotropic glutamate receptors. *Neuron* **69**, 44–60 (2011).
- van der Goes van Naters, W. & Carlson, J. R. Receptors and neurons for fly odors in *Drosophila*. *Curr. Biol.* **17**, 606–612 (2007).
- Ferveur, J. F. Cuticular hydrocarbons: their evolution and roles in *Drosophila* pheromonal communication. *Behav. Genet.* **35**, 279–295 (2005).
- Yao, C. A., Ignell, R. & Carlson, J. R. Chemosensory coding by neurons in the coeloconic sensilla of the *Drosophila* antenna. *J. Neurosci.* **25**, 8359–8367 (2005).

- Barata, A. *et al.* Analytical and sensorial characterization of the aroma of wines produced with sour rotten grapes using GC-O and GC-MS: identification of key aroma compounds. *J. Agric. Food Chem.* **59**, 2543–2553 (2011).
- Kim, J., Jeon, C. O. & Park, W. A green fluorescent protein-based whole-cell bioreporter for the detection of phenylacetic acid. *J. Microbiol. Biotechnol.* **17**, 1727–1732 (2007).
- Bray, S. & Amrein, H. A putative *Drosophila* pheromone receptor expressed in male-specific taste neurons is required for efficient courtship. *Neuron* **39**, 1019–1029 (2003).
- Kohatsu, S., Koganezawa, M. & Yamamoto, D. Female contact activates male-specific interneurons that trigger stereotypic courtship behavior in *Drosophila*. *Neuron* **69**, 498–508 (2011).
- Kurtovic, A., Widmer, A. & Dickson, B. J. A single class of olfactory neurons mediates behavioural responses to a *Drosophila* sex pheromone. *Nature* **446**, 542–546 (2007).
- Root, C. M. *et al.* A presynaptic gain control mechanism fine-tunes olfactory behavior. *Neuron* **59**, 311–321 (2008).
- Wang, L. *et al.* Hierarchical chemosensory regulation of male–male social interactions in *Drosophila*. *Nature Neurosci.* **14**, 757–762 (2011).
- Vosshall, L. B. & Stocker, R. F. Molecular architecture of smell and taste in *Drosophila*. *Annu. Rev. Neurosci.* **30**, 505–533 (2007).
- Jefferis, G. S. *et al.* Comprehensive maps of *Drosophila* higher olfactory centers: spatially segregated fruit and pheromone representation. *Cell* **128**, 1187–1203 (2007).
- Cachero, S., Ostrovsky, A. D., Yu, J. Y., Dickson, B. J. & Jefferis, G. S. Sexual dimorphism in the fly brain. *Curr. Biol.* **20**, 1589–1601 (2010).
- Yu, J. Y., Kanai, M. I., Demir, E., Jefferis, G. S. & Dickson, B. J. Cellular organization of the neural circuit that drives *Drosophila* courtship behavior. *Curr. Biol.* **20**, 1602–1614 (2010).
- Croset, V. *et al.* Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction. *PLoS Genet.* **6**, e1001064 (2010).
- Ewing, A. W. & Manning, A. The effect of exogenous scent on the mating of *Drosophila melanogaster*. *Anim. Behav.* **11**, 596–598 (1963).
- Shorey, H. H. & Bartell, R. J. Role of a volatile female sex pheromone in stimulating male courtship behaviour in *Drosophila melanogaster*. *Anim. Behav.* **18**, 159–164 (1970).
- Spieth, H. T. Courtship behavior in *Drosophila*. *Annu. Rev. Entomol.* **19**, 385–405 (1974).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to B. Dickson, A. Hofbauer, T. Lee, the Bloomington *Drosophila* Stock Center, the *Drosophila* Species Stock Center and the Developmental Studies Hybridoma Bank for provision of plasmid vectors, *Drosophila* strains and antibodies. We are also grateful to A. Wong, J. Wang, R. Axel, H.-H. Yu and T. Lee for sharing raw image data. We thank D. Featherstone, J.-F. Ferveur, T. Kaweck, L. Keller, S. Martin and members of the Benton laboratory for comments on the manuscript. Y.G., J.-P.F. and J.C. are supported by the Centre National de la Recherche Scientifique (CNRS), the Agence Nationale de la Recherche (ANR; JCJC, GGCB-2010) and the Conseil Régional de Bourgogne (FABER). R.R. was supported by a Roche Research Foundation fellowship. G.S.X.E.J. is supported by the Medical Research Council and a European Research Council Starting Investigator Grant. Research in R.B.'s laboratory is supported by the University of Lausanne, a European Research Council Starting Independent Researcher Grant and the Swiss National Science Foundation.

Author Contributions Y.G. and R.B. conceived the project. Y.G. performed the gene-targeting screen, contributed to the histological analysis and performed most of the behavioural experiments. R.R. performed the electrophysiological odour response screen, the fly odour stimulation assays and the phylogenetic analyses. J.-P.F. performed the chemical analysis. L.A. assisted in the generation and characterization of transgenic flies and contributed to the histological analysis. J.C. contributed to the behavioural experiments. G.S.X.E.J. performed the analysis of projection neurons. R.B. generated the DNA constructs, performed all other electrophysiological analyses and wrote the paper with contributions from Y.G., R.R., J.-P.F. and G.S.X.E.J.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.B. (Richard.Benton@unil.ch).

METHODS

Drosophila strains. *Drosophila* stocks were maintained on a standard corn flour, yeast and agar medium under a 12 h light and 12 h dark cycle at 25 °C. We used the following mutant and transgenic strains: *Or35a-GAL4* (ref. 13), *Ir84a-GAL4* (ref. 9), *UAS-Ir84a*⁶, *UAS-EGFP-Ir84a*¹⁰, *UAS-mCD8:GFP*⁹, *fru*^{GAL4} (ref. 5), *fruP1-GAL4* (ref. 4), *lexAop-rCD2::GFP*, *UAS-mCD8* (ref. 30), *UAS-DTP*¹, *70FLP, 70I-Scel/Cyo*, *70FLP* and *70I-Cre*¹². For behavioural analysis, all flies carrying mutant and transgenic chromosomes were backcrossed to an isogenized *w*¹¹¹⁸ control line for five generations. The sequenced wild-type *D. melanogaster* strain (Bloomington *Drosophila* Stock Center number 2057) was used for the ac4 odour screen, and Canton-S was used as the courtship object in behavioural experiments. Wild-type *D. mojavensis* was obtained from the *Drosophila* Species Stock Center (University of California, San Diego; stock number 15081-1352.22).

Molecular biology. New transgenes were constructed as follows, using standard methods. For the *Ir84a* *GAL4* targeting construct, 5,087 base pairs (bp) genomic DNA immediately upstream of the *Ir84a* start codon was amplified and subcloned as 1,924-bp and 3,163-bp fragments either side of the I-SceI site upstream of *GAL4* in the pED36 targeting vector³. Similarly, a 964-bp genomic fragment immediately downstream of the *Ir84a* stop codon was cloned downstream of *GAL4*. For the *Ir84a* promoter-*lexA* construct, the same 1,964 bp *Ir84a* promoter fragment used in the *Ir84a-GAL4* transgene⁹ was cloned upstream of *lexA:VP16-SV40* (ref. 30) in the vector pattB³³.

Gene targeting. Targeted replacement of *Ir84a* by *GAL4* was performed essentially as described previously¹⁸. A transgenic line containing the *Ir84a* *GAL4* targeting construct on the X chromosome was used as a donor for targeting crosses in which heat-shock-inducible transgenic FLP and I-SceI enzymes were used to excise and linearize the targeting fragment. Reinsertion was detected by the presence of a FLP-insensitive *white*⁺ marker. From ~193,000 flies screened, five targeted insertions were recovered and verified by PCR as duplicates of the *Ir84a* locus. One duplicate was resolved with I-CreI, selecting for loss of *white*⁺. From 218 *white*⁻ flies, 1 was found to contain the desired precise *Ir84a* to *GAL4* replacement.

Histology. Immunofluorescence on antennal cryosections was performed as described previously^{10,34}. Immunofluorescence on whole mount brains, and combined fluorescent RNA *in situ* hybridization and immunofluorescence on antennal sections, was performed as described previously⁹. The primary antibodies used were mouse monoclonal nc82 (1:10 dilution; provided by A. Hofbauer), mouse monoclonal 21A6 (1:200; Developmental Studies Hybridoma Bank), rabbit anti-GFP (1:1,000; Invitrogen), mouse anti-GFP (1:500; Invitrogen), rat anti-mCD8 (1:50; Caltag), anti-DIG-POD (1:500; Roche). The secondary antibodies used were Alexa Fluor 488-, Cy3- and Alexa Fluor 647-conjugated goat anti-mouse, anti-rabbit or anti-rat IgG (Molecular Probes and Jackson ImmunoResearch). These were diluted to 1:100 and 1:1,000 for whole mount brains and antennal sections, respectively. All microscopy was performed using an LSM 510 laser scanning confocal microscope (Zeiss).

Electrophysiology. Extracellular recordings in single sensilla of 1–8-day-old flies were performed and quantified as described previously⁶. Odours were diluted to 1% (v/v) unless otherwise noted. The CAS numbers, solvents, sources and purities for the odorants used are provided in Supplementary Table 2. Corrected responses were quantified by counting all spikes in ac4 sensilla in a 0.5 s window from 150–200 ms after the stimulus trigger. This delay period was defined precisely for each recording session with a control odour response and mainly represents the time for the odour to reach the preparation. From this spike value, we subtracted the number of spontaneous spikes in a 0.5 s window before stimulation and doubled the result to obtain spikes s⁻¹. Spontaneous activity was quantified by counting the spikes in a 5 s window without stimulus and then dividing by five to obtain spikes s⁻¹. After verifying that the responses were normally distributed, we compared all genotypes for a given experiment by analysis of variance (ANOVA), with the genotype as the main effect, and adjusted the alpha level for planned post hoc means comparisons.

To assess the responses of live flies to volatile compounds, ~100–120 male or virgin female flies were placed in a 5 ml glass pipette (witeg Labortechnik). The pipette was then closed at each end with cotton gauze, and an air pulse (3 s) was blown through it. As controls, an empty pipette was used or a pipette loaded with 10 µl phenylacetic acid (0.1 µg µl⁻¹ in paraffin oil, as 1 µl drops on ten individual 2 mm² filter papers). The responses were quantified by counting all of the spikes in a 1 s window from 500 ms after the odour stimulus trigger and then subtracting the number of spontaneous spikes in a 1 s window before stimulation. To assess the responses to *D. melanogaster* cuticular hydrocarbons, we prepared extracts in 50 µl CH₂Cl₂ as described below (except without the *n*-C13 standard) and then evaporated the solvent under N₂ and diluted the extract in paraffin oil to a concentration of 10 fly equivalents µl⁻¹. These extracts (1 µl) (or control stimuli) were placed on the tip of a borosilicate glass capillary (World Precision Instruments) and brought

into close proximity (~5 mm) to the antenna. The changes in spiking frequency were quantified by measuring the spontaneous spiking frequency in spikes s⁻¹ in a 3 s window before the stimulus was brought close to the antenna and then subtracting this value from the spiking frequency in a 3 s window when the stimulus was placed close to the antenna. This second 3 s window typically started 3–5 s after stimulus placement, a time frame when the responses to phenylacetic acid were most robust.

Extract preparation and chemical analysis. As phenylacetaldehyde and phenylacetic acid are polar compounds, we used CH₂Cl₂ for all extractions. *D. melanogaster* extracts were prepared by incubating five to ten 5-day-old flies for 5 min in vials containing 50 µl CH₂Cl₂ and 100 ng *n*-C13 (as an internal standard) at room temperature. After the flies were removed, the extracts were stored at -20 °C until analysis. For standard fly food and fruit extracts, 2 g material was crushed in 20 ml vials containing 4 ml CH₂Cl₂ and 200 ng *n*-C13 as an internal standard. After vortexing for 10 min at room temperature, the extracts were filtered through glass wool and concentrated to ~100 µl under a gentle flow of N₂. The extracts were stored at -20 °C until analysis.

Because of the difficulty of detecting phenylacetic acid and phenylacetaldehyde at low concentration in the extracts, we tested three column types with different polarities: a CP-Sil 5 CB (apolar type, 25 m × 0.25 mm internal diameter, 0.12 µm film thickness; Varian), a CP-Wax 58 CB (polar type, 25 m × 0.25 mm internal diameter, 0.20 µm film thickness; Varian) and a VF-1ms (medium polarity type, 20 m × 0.15 mm internal diameter, 0.15 µm film thickness; Varian). Only the last two columns gave satisfactory quantitative results, and these were used for all subsequent analyses. The extracts were analysed using a GC-MS-QP2010 apparatus (Shimadzu) in splitless mode, fitted with a VF-1ms (*D. melanogaster* extracts) or a CP-Wax 58 CB (standard fly food, banana and prickly-pear extracts) fused silica capillary column. The columns were held isothermally at 40 °C for 2 min, then the temperature was programmed to increase at a rate of 3 °C min⁻¹ to 240 °C (CP-Wax 58CB) or 300 °C (VF-1ms). Helium was used as the carrier gas at a linear velocity of 47 cm s⁻¹. The injector port was set at 280 °C. The mass spectrometer was operated at 70 eV, and scanning was performed from 29 to 600 AMU at 0.5 scans s⁻¹. The injection split was opened 1 min after the injection. Phenylacetic acid and phenylacetaldehyde were identified using their retention time and their fragmentation patterns; diagnostic ions were compared with both the NIST/EPA/NIH Mass Spectral Library and the mass spectrum of synthetic chemical standards (Sigma-Aldrich) analysed under the same conditions. For quantitative analyses of phenylacetic acid and phenylacetaldehyde, the response factor was determined for each of the compounds and *n*-C13 at 0.1, 0.5, 1, 5 and 10 ng on both types of column. For analysis of extracts from flies grown on minimal medium, late second or young third instar larvae reared on standard food medium were washed six times in distilled water to remove any traces of medium. Larvae were then transferred to a minimal growth medium (0.1 M sucrose and 1% agarose). At emergence, adults were separated by sex and aged in groups of 20 flies on minimal growth medium for 5 days, before extraction as described above.

Courtship assays. Male and virgin female flies were collected at eclosion. They were kept in vials, individually for subjects or in groups of 10 for objects, for 5–9 days before behavioural assays. The vials contained fresh food and were kept under a 12 h light and 12 h dark cycle at 25 °C. Behavioural assays were performed at approximately the same time each morning. Single-pair courtship tests were performed in round observation chambers (1.2 cm in diameter and 0.4 cm deep, with nine-chamber plexiglass courtship wheels) under far-red light (to eliminate the contribution of visual cues) at 25 °C. For most experiments, object females (or males) were decapitated with a clean razor blade (to avoid any reciprocal interaction during courtship) and were left to 'recover' for ~10 min before introduction into the courtship chamber. We also tested males for their courtship behaviour in more natural conditions: using intact live virgin Canton-S females, together with ~90 mg fly food medium (corresponding to approximately one-quarter of the courtship chamber volume). Male courtship behaviour towards these different types of object was recorded for 10 min with an HDR-SR10 digital video camera (Sony). After recording, videos were analysed by a researcher who was blinded to the genotypes, using Annotation software (SaySoSoft). The courtship index was calculated as the percentage of time that a male courted the object during a 10-min period. Female copulation behaviour was quantified by direct observation for 1 h without video recording: copulation latency represents the time to the beginning of copulation from the moment that the virgin female and the male are placed in the chamber; copulation success is the percentage of flies that copulate; and copulation duration is the total time of copulation from the start to the end of copulation.

For perfumed assays, the courtship chamber was placed on top of an identical chamber, and the chambers were separated by muslin gauze to avoid direct contact between the flies in the top chamber and the chemical stimuli in the bottom chamber (food, 50 ± 2 mg standard fly food; solvent, 10 µl paraffin oil (on a filter paper or aluminium foil pad); phenylacetic acid, 10 µl ~35 µg µl⁻¹ solution in

paraffin oil (that is, $\sim 2.5 \mu\text{mol}$ total odour); ethyl butyrate, $10 \mu\text{l}$ 1% (v/v) solution in paraffin oil (that is, $\sim 0.8 \mu\text{mol}$ total odour)). Single object males were placed in the upper chamber with a killed intact virgin wild-type female at 25°C . The female had been anaesthetized with CO_2 , placed in an Eppendorf tube and killed by submerging the tube in liquid nitrogen. The female was then warmed for ~ 10 min at 25°C in a Petri dish before being placed in the courtship chamber. Males court killed objects only at low levels, facilitating the determination of the enhancing effect of perfuming.

Olfactory assays. Male olfactory responses were tested using an adapted Y-maze assay³⁵; in brief, three-way plastic tube connectors (Reactolab) were used to join three glass vials by way of 1 ml pipette tips that pierced the cotton vial tops, forming a tightly sealed Y maze. The narrow ends of the pipette tips were cut and oriented in each vial top to form one 'loading' vial and two 'trap' vials, which contained $40 \mu\text{l}$ odour (10% (v/v) acetic acid in water (that is, $\sim 70 \mu\text{mol}$ total odour) or $\sim 35 \mu\text{g} \mu\text{l}^{-1}$ phenylacetic acid in paraffin oil (that is, $\sim 10 \mu\text{mol}$ total odour)) or $40 \mu\text{l}$ corresponding solvent on filter paper. Ten 5–9-day-old males (starved for 18–21 h in glass tubes with water on filter paper) were introduced without CO_2 anaesthesia into the loading vial and were allowed 2 h at 25°C to choose to enter the trap vial containing the odour or the solvent. The resultant olfactory index was obtained with the following formula: (number in the odour tube – number in the solvent tube)/total number of loaded males.

Projection neuron analysis. Projection neuron images (275) from four different sources were brought into a common reference space for analysis^{22,36–38}. The common template (Cell07) was from ref. 22, so 236 projection neuron images from that study did not require further processing. The anti-Discs large channel of 34 confocal images from ref. 38 was registered to a single female template brain (FlyCircuit ID 6475, FC6475), which was chosen for good quality staining and imaging, using the CMTK toolkit (<http://www.nitrc.org/projects/cmtk>) as described previously²³. The FC6475 template brain was then registered to the Cell07 brain using landmarks registration based on 24 manually chosen anatomical landmarks visible in both anti-Discs large and nc82 staining. This landmarks registration used the pnreg command of the IRTK toolkit (<http://www.doc.ic.ac.uk/~dr/software/>), which uses nonlinear third-order B-spline registration³⁹. The fiducial registration error, measured as the root mean squared deviation, was $5.1 \mu\text{m}$ after affine registration and $2.4 \mu\text{m}$ after warping registration, corresponding to per axis accuracies of 2.9 and $1.4 \mu\text{m}$, respectively. The final per axis registration accuracy for independent landmarks (not used during the registration) was $3.3 \mu\text{m}$, which is comparable to that of previous studies^{22,24}.

Three VL2a projection neuron images from ref. 37 were registered to the Cell07 template by choosing 12–14 landmarks in the nc82 channel present in each stack. Landmarks registration used the pnreg command of IRTK. Two VL2a projection neuron images from ref. 36 were registered to the Cell07 template using 10–14 landmarks. Although those brains had weak nc82 staining, the only reliable marker was GH146-GAL4-driven expression of a CD2 reporter. We therefore used a two-channel confocal image of an nc82-stained brain with GH146-driven mCD8:GFP expression that had been registered by way of its nc82 channel to the Cell07 template to help choose landmarks. Landmarks registration again used the pnreg command of IRTK. The fiducial registration error (root mean squared deviation) for the five brains varied between 3.6 and $4.9 \mu\text{m}$ (affine) and 1.8 and $3.1 \mu\text{m}$ (warp).

Neuronal tracing was carried out in Amira (<http://www.amira.com>) with the hxskeletonize plug-in⁴⁰. The three-dimensional coordinates of neuronal tracings were then transformed from their respective original image coordinates using the CMTK gregxform tool (FlyCircuit images \rightarrow FC6475 template) and the IRTK ptransform tool (FC6475 \rightarrow Cell07, three VL2a images \rightarrow FC6475). The underlying tools were called using custom code written in R (<http://www.r-project.org>), which was also used to analyse the transformed neuronal tracings (using the AnalysisSuite R package)²². The final visualization was carried out in Amira using Tcl scripts to load, show/hide and colour neuron tracings and to take snapshots. R and Tcl source code, landmark files, template brain images and neuronal tracings will be available on publication at the Jefferis laboratory website (<http://flybrain.mrc-lmb.cam.ac.uk>).

Analysis of axonal overlap. The degree of overlap between the axon terminals of different projection neurons was measured by making use of the three-dimensional convex hull of each terminal (using the Qhull library⁴¹ as exposed by the R geometry package). It may help to note that for a set of two-dimensional points in the plane, the convex hull can be obtained by stretching an elastic band to encompass the whole object and then releasing it. To obtain an overlap score for a projection

neuron pair (A,B), we calculated three convex hulls, $H(A)$, $H(B)$ and $H(A,B)$, where $H(A,B)$ is the convex hull for all of the points in A and B. We then calculated a normalized overlap score as $s(A,B) = (H(A) + H(B) - H(A,B))/(H(A) + H(B))$.

If A and B are identical (that is, complete overlap), the score will be 0.5. Less overlap will result in lower scores, with negative scores when the amount of intervening space between the axon terminals exceeds the overlap (if any). We first calculated an overlap score for all 275 neurons in the data set. We then aggregated the scores for all pairwise combinations of the 46 projection neuron classes in the data set and calculated the median score for each combination. This resulted in a symmetrical distance matrix with $46 \times 45/2 = 1,035$ unique off-diagonal entries. The median overlap score for VL2a and VA11m projection neurons was 0.224, which was at the 97.7th centile: that is, the overlap between these two classes was among the strongest 2.3% in the data set. VL2a was the neuronal class with the strongest overlap with VA11m (and vice versa). Cluster analysis of these data was performed with Ward's method, using the hclust function of R. Overlap scores, s , were converted to a distance, d , suitable for clustering by the simple transform $d = 0.5 - s$.

The VL2a projection neuron image data were obtained in separate studies from the bulk of the projection neuron data, so it is natural to ask whether this could have had some effect on our analysis. However, the consistent results for VL2a projection neuron images obtained from two different laboratories argue very strongly against this possibility. Visual inspection indicated that there was excellent overlap between the VL2a neurons from the two sources, and the pairwise overlap scores between and within the two groups were almost identical (median 0.284 and median 0.301, corresponding to the 98.0th and 98.8th centiles in Supplementary Fig. 4b). As an additional check, we repeated the axon overlap analysis and clustering with affine registered VL2a neurons (affine registration is more robust, although it is less accurate in the ideal situation). This had a negligible effect on the VL2a/VA11m median overlap score (0.197 and 0.224, corresponding to the 96.2th centile and 97.7th centile, respectively) and had no effect on the clustering results (G.S.X.E.J., data not shown).

Phylogenetic analysis. The phylogenetic tree of selected antennal-expressed IR orthologous groups²⁵ was constructed using PhyML⁴² and is based on the most conserved columns of an amino acid alignment generated with MUSCLE⁴³. The phylogeny was rooted using IR25a. iTOL⁴⁴ was used to view and graphically edit the tree.

29. Lee, T. & Luo, L. Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. *Neuron* **22**, 451–461 (1999).
30. Lai, S. L. & Lee, T. Genetic mosaic with dual binary transcriptional systems in *Drosophila*. *Nature Neurosci.* **9**, 703–709 (2006).
31. Han, D. D., Stein, D. & Stevens, L. M. Investigating the function of follicular subpopulations during *Drosophila* oogenesis through hormone-dependent enhancer-targeted cell ablation. *Development* **127**, 573–583 (2000).
32. Rong, Y. S. et al. Targeted mutagenesis by homologous recombination in *D. melanogaster*. *Genes Dev.* **16**, 1568–1581 (2002).
33. Bischof, J., Maeda, R. K., Hediger, M., Karch, F. & Basler, K. An optimized transgenesis system for *Drosophila* using germ-line-specific ϕC31 integrases. *Proc. Natl Acad. Sci. USA* **104**, 3312–3317 (2007).
34. Benton, R., Sachse, S., Michnick, S. W. & Vossahl, L. B. Atypical membrane topology and heteromeric function of *Drosophila* odorant receptors *in vivo*. *PLoS Biol.* **4**, e20 (2006).
35. Alcorta, E. & Rubio, J. Intrapopulation variation of olfactory responses in *Drosophila melanogaster*. *Behav. Genet.* **19**, 285–299 (1989).
36. Wong, A. M., Wang, J. W. & Axel, R. Spatial representation of the glomerular map in the *Drosophila* protocerebrum. *Cell* **109**, 229–241 (2002).
37. Yu, H. H. et al. A complete developmental sequence of a *Drosophila* neuronal lineage as revealed by twin-spot MARCM. *PLoS Biol.* **8**, e1000461 (2010).
38. Chiang, A. S. et al. Three-dimensional reconstruction of brain-wide wiring networks in *Drosophila* at single-cell resolution. *Curr. Biol.* **21**, 1–11 (2011).
39. Rueckert, D. et al. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans. Med. Imaging* **18**, 712–721 (1999).
40. Evers, J. F., Schmitt, S., Sibila, M. & Duch, C. Progress in functional neuroanatomy: precise automatic geometric reconstruction of neuronal morphology from confocal image stacks. *J. Neurophysiol.* **93**, 2331–2342 (2005).
41. Barber, C. B., Dobkin, D. P. & Huhdanpaa, H. The Quickhull algorithm for convex hulls. *ACM Trans. Math. Softw.* **22**, 469–483 (1996).
42. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704 (2003).
43. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
44. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**, 127–128 (2007).

STING is a direct innate immune sensor of cyclic di-GMP

Dara L. Burdette¹, Kathryn M. Monroe¹, Katia Sotelo-Troha¹, Jeff S. Iwig^{1,2}, Barbara Eckert¹, Mamoru Hyodo³, Yoshihiro Hayakawa⁴ & Russell E. Vance¹

The innate immune system detects infection by using germline-encoded receptors that are specific for conserved microbial molecules. The recognition of microbial ligands leads to the production of cytokines, such as type I interferons (IFNs), that are essential for successful pathogen elimination. Cytosolic detection of pathogen-derived DNA is one major mechanism of inducing IFN production^{1,2}, and this process requires signalling through TANK binding kinase 1 (TBK1) and its downstream transcription factor, IFN-regulatory factor 3 (IRF3). In addition, a transmembrane protein called STING (stimulator of IFN genes; also known as MITA, ERIS, MPYS and TMEM173) functions as an essential signalling adaptor, linking the cytosolic detection of DNA to the TBK1–IRF3 signalling axis^{3–7}. Recently, unique nucleic acids called cyclic dinucleotides, which function as conserved signalling molecules in bacteria⁸, have also been shown to induce a STING-dependent type I IFN response^{9–12}. However, a mammalian sensor of cyclic dinucleotides has not been identified. Here we report evidence that STING itself is an innate immune sensor of cyclic dinucleotides. We demonstrate that STING binds directly to radiolabelled cyclic diguanylate monophosphate (c-di-GMP), and we show that unlabelled cyclic dinucleotides, but not other nucleotides or nucleic acids, compete with c-di-GMP for binding to STING. Furthermore, we identify mutations in STING that selectively affect the response to cyclic dinucleotides without affecting the response to DNA. Thus, STING seems to function as a direct sensor of cyclic dinucleotides, in addition to its established role as a signalling adaptor in the IFN response to cytosolic DNA. Cyclic dinucleotides have shown promise as novel vaccine adjuvants and immunotherapeutics^{9,13}, and our results provide insight into the mechanism by which cyclic dinucleotides are sensed by the innate immune system.

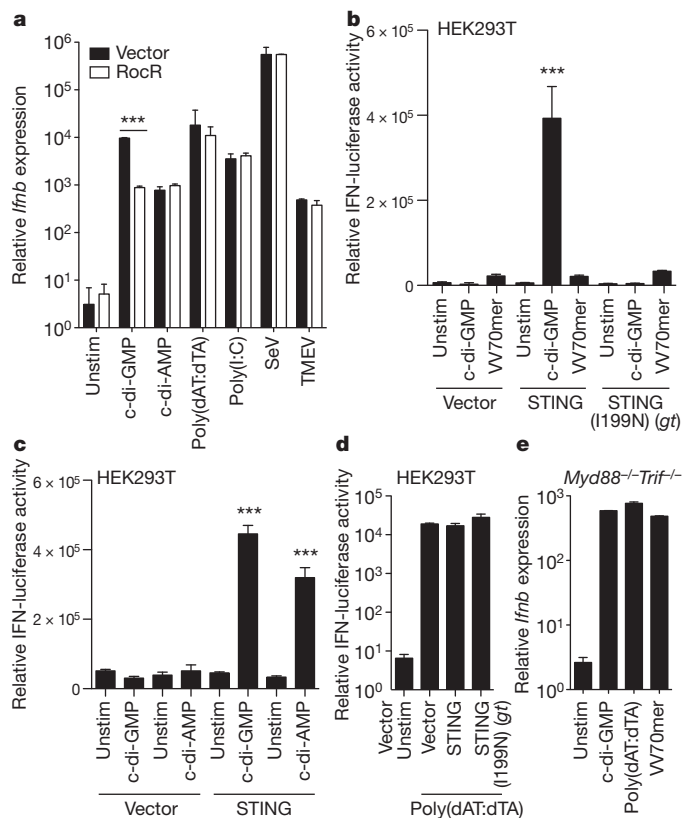
Nucleotides are crucial signalling molecules in all domains of life, but cyclic dinucleotides seem to be produced solely by Bacteria and Archaea. For example, c-di-GMP is a ubiquitous second messenger that regulates biofilm formation, motility and virulence in a diverse range of bacterial species⁸. Recently, cyclic diadenylate monophosphate (c-di-AMP) was discovered to be a bacterial regulatory molecule¹⁴, although its role remains to be fully characterized. Because they are unique to microorganisms, cyclic dinucleotides are appropriate targets for immune recognition¹⁵. Indeed, the induction of IFN production by *Listeria monocytogenes* depends on bacterial secretion of cyclic-di-AMP¹². However, it remains unclear how cyclic dinucleotides are sensed in mammalian cells.

To address the mechanism by which mammalian cells sense cyclic dinucleotides, we first confirmed that cyclic dinucleotides are detected in the host cell cytosol¹⁰ by expressing RocR, a c-di-GMP-specific phosphodiesterase from *Pseudomonas aeruginosa*, in the cytosol of macrophages. In these cells, IFN induction in response to c-di-GMP (but not other stimuli) is tenfold lower than that in control, vector-transduced, cells (Fig. 1a), confirming that the cytosolic presence of c-di-GMP is important for inducing IFN.

To identify candidate cyclic dinucleotide sensors, we sought to identify molecules that could reconstitute the IFN response to cyclic dinucleotides in HEK293T cells, which do not respond to c-di-GMP¹⁰. Because STING is essential for the IFN response to cyclic dinucleotides¹¹ and because STING expression is low or undetectable in HEK293T cells (Supplementary Fig. 1 and data not shown), we first expressed STING in HEK293T cells. The overexpression of STING spontaneously induces an IFN reporter^{3,6,7}, so we transfected a small amount of *Sting*-encoding vector that by itself was insufficient to induce IFN. Low levels of STING protein were sufficient to reconstitute the responsiveness of HEK293T cells to c-di-GMP (Fig. 1b) and c-di-AMP (Fig. 1c). By contrast, the non-functional goldenticket (*gt*) allele of *Sting* (which results in a STING protein in which asparagine has been substituted for isoleucine, I199N)¹¹ did not restore responsiveness to c-di-GMP (Fig. 1b). Interestingly, the expression of wild-type STING did not reconstitute the responsiveness of HEK293T cells to double-stranded DNA (dsDNA) oligonucleotides (for example, a 70-base-pair oligonucleotide from vaccinia virus (VV70mer) or IFN-stimulatory DNA (ISD)) that had previously been shown to induce type I IFNs in macrophages through STING^{4,16} (Fig. 1b and Supplementary Fig. 2a). By contrast, the induction of IFN by poly(dA:dT)·poly(dT:dA) DNA (denoted poly(dAT:dTA)) was identical in cells that were transfected with wild-type *Sting* and those transfected with *gt*, demonstrating that the RNA polymerase III DNA-sensing pathway^{17,18} is intact in these cells and is not responsible for the detection of c-di-GMP (Fig. 1d). As a positive control, *Myd88*^{−/−}*Trif*^{−/−} immortalized macrophages, which express STING, responded similarly to c-di-GMP, poly(dAT:dTA), VV70mer and ISD (Fig. 1e and Supplementary Fig. 2b). Together, our results show that STING expression is sufficient to restore the responsiveness of HEK293T cells to cyclic dinucleotides but not to DNA.

We next tested whether STING, or perhaps another protein in HEK293T cells, binds to c-di-GMP. We used HEK293T cell lysates in an *in vitro* ultraviolet radiation crosslinking assay to identify putative sensor proteins that interact directly with radiolabelled c-di-GMP (c-di-[³²P]GMP). We expected to identify directly interacting proteins because only molecules within bond-length proximity are efficiently crosslinked by ultraviolet radiation¹⁹. We detected a prominent ~40-kDa radiolabelled protein, which corresponds to the predicted molecular weight of monomeric STING, in the lysates of cells transfected with a vector encoding haemagglutinin (HA)-tagged STING (STING–HA) but not in the lysates of cells transfected with a vector encoding STING(I199N)–HA or vector only (Fig. 2a). The ~40-kDa band did not appear when the same lysates were crosslinked in the presence of [³²P]GTP, implying that crosslinking to c-di-[³²P]GMP was specific (Fig. 2a). We also observed an ~80-kDa species, which might correspond to a previously reported STING dimer⁶ (Fig. 2b). To test the hypothesis that STING crosslinks with c-di-[³²P]GMP, we immunoprecipitated STING from transfected HEK293T cells and performed the c-di-[³²P]GMP crosslinking assay on the

¹Department of Molecular & Cell Biology, University of California, Berkeley, California 94720, USA. ²Department of Chemistry, Howard Hughes Medical Institute, University of California, Berkeley, California 94720, USA. ³Faculty of Pharmaceutical Science, Hokkaido University, Kita 12, Nishi 6, Kita-ku, Sapporo, Hokkaido 060-0812, Japan. ⁴Department of Applied Chemistry, Faculty of Engineering, Aichi Institute of Technology, 1247 Yachigusa, Yakusa, Toyota 470-0392, Japan.



immunoprecipitates. Bands corresponding to the molecular weight of the STING monomer and dimer were identified only in immunoprecipitates from lysates overexpressing STING and not in mock immunoprecipitates from lysates of vector-only-transfected cells (Fig. 2b). Thus, STING seems to bind to c-di-GMP.

To confirm that the binding of c-di-[32 P]GMP to STING is specific, we performed the c-di-GMP crosslinking assay in the presence of

Figure 1 | STING is sufficient to restore responsiveness to cyclic dinucleotides. **a**, Immortalized *Myd88*^{-/-}*Trif*^{-/-} macrophages were transduced with retrovirus expressing *rocR* and then stimulated for 6 h. IFN induction was measured by quantitative PCR with reverse transcription (quantitative RT-PCR) and normalized to ribosomal protein 17 (*Rps17*) expression levels. **b–d**, HEK293T cells were transfected as indicated, together with an IFN-luciferase reporter, and luciferase activity was measured 6 h after stimulation. **e**, *Myd88*^{-/-}*Trif*^{-/-} macrophages were stimulated for 6 h, and IFN induction was measured as in **a**. **a–e**, Data are presented as mean \pm s.d. ($n = 3$) and are representative of at least three independent experiments. ***, $P < 0.001$. SeV, Sendai virus; TMEV, Theiler's murine encephalomyelitis virus; unstim, unstimulated; VV70mer, a stimulatory dsDNA oligonucleotide derived from vaccinia virus.

unlabelled nucleotides. Unlabelled c-di-GMP and c-di-AMP specifically competed with c-di-[32 P]GMP for binding to STING (Fig. 2c, d). By contrast, GTP, other guanosine derivatives and nucleic acids (including dsDNA) competed away nonspecific binding (Fig. 2c, d, asterisk); however, under our specific assay conditions, these molecules could not compete efficiently with c-di-[32 P]GMP for binding to STING (Fig. 2c, d, arrow). Because the cell cytosol contains high concentrations of GTP (0.1–1 mM), a putative c-di-GMP sensor must have a high degree of specificity for c-di-GMP over GTP. We found that c-di-GMP efficiently crosslinked to STING even in the presence of 1 mM GTP (Fig. 2c).

Although these data imply that STING directly and specifically binds to cyclic dinucleotides, they do not address whether other host proteins might also be required. STING is predicted to encode an amino-terminal domain with multiple transmembrane segments, followed by a globular carboxy-terminal domain (CTD). Because the CTD contains the amino acid substitution that abolishes STING function in *gt* mice (I199N)¹¹, we suspected that the CTD might be involved in binding to cyclic dinucleotides. Thus, we subjected purified recombinant His₆-tagged STING CTD (amino acids 138–378) (Fig. 2e) to the c-di-[32 P]GMP crosslinking assay. We found that the recombinant CTD of STING bound to c-di-[32 P]GMP and that this binding was specifically competed away with cold (unlabelled) c-di-GMP or c-di-AMP but not with cold GTP or ATP (Fig. 2f). We used equilibrium

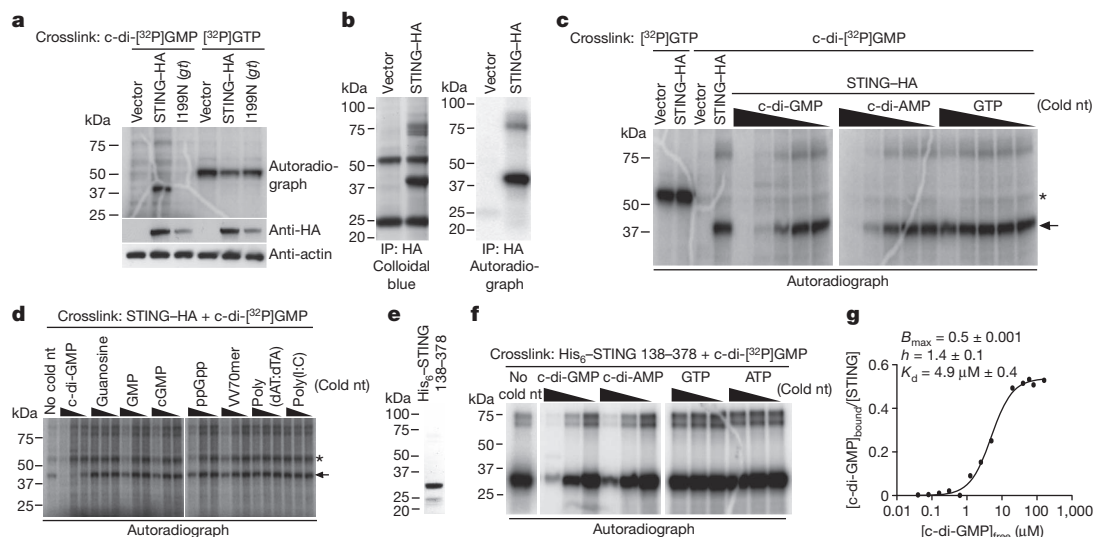


Figure 2 | STING binds cyclic dinucleotides. **a**, HEK293T cells were transfected as indicated, and the cell lysates were subjected to an *in vitro* ultraviolet radiation crosslinking assay with c-di-[32 P]GMP. Samples were separated by SDS-PAGE and visualized by autoradiography or western blotting. **b**, HEK293T cells were transfected as in **a**, and anti-HA immunoprecipitates were stained with colloidal blue or subjected to the crosslinking assay with c-di-[32 P]GMP. **c**, **d**, HEK293T cells were transfected as in **a**, and the cell lysates were crosslinked with ultraviolet radiation to c-di-[32 P]GMP or [32 P]GTP in the presence of cold competing nucleotides in tenfold serial dilutions beginning at 1 mM (as indicated by the wedges), except

for guanosine (beginning at 0.1 mM), VV70mer (500 μ g ml⁻¹), poly(dAT:dTA) (50 μ g ml⁻¹) and poly(I)•poly(C) (denoted poly(I:C)) (50 μ g ml⁻¹). The arrow indicates STING, and the asterisk indicates a nonspecific band. nt, nucleotides; ppGpp, guanosine-3',5'-bisphosphate. **e**, His₆-STING (amino acids 138–378) (1 μ g) was separated by SDS-PAGE and stained with Coomassie blue. **f**, His₆-STING 138–378 was analysed as in **c**. **g**, The binding of c-di-GMP to purified His₆-STING (10 μ M) was measured by equilibrium dialysis. B_{\max} , maximum number of binding sites; h , hill coefficient; K_d , dissociation constant. **a–g**, Data are representative of three independent experiments.

dialysis to obtain an estimate of the affinity (dissociation constant, K_d) of c-di-GMP binding to the STING CTD, which was $\sim 5 \mu\text{M}$ (Fig. 2g). In its native membrane-bound form, or in complex with other host factors, STING may have a stronger affinity for c-di-GMP; nevertheless, a $5 \mu\text{M}$ affinity is consistent with the dose response that has previously been observed in macrophages¹². Consistent with the ability of STING to dimerize⁶, the binding data suggest a stoichiometry of one molecule of c-di-GMP per two molecules of STING.

To identify the amino acids involved in c-di-GMP binding and/or IFN induction, we introduced point mutations into STING. Focusing on clusters of conserved and charged residues, we mutated 67 amino acids, either individually or in groups, and we classified these mutants into five categories (Fig. 3, Supplementary Table 1 and Supplementary Figs 3 and 4). Class I consists of mutants that abolish both binding and IFN induction (Fig. 3a–c, red, and Supplementary Table 1). Class II mutants bind to c-di-GMP but fail to induce IFN (Fig. 3c, purple). Class III comprises ‘hyperactive’ mutants, which spontaneously induce IFN at low levels of transfection (Fig. 3a–c, green, and Supplementary Table 1). Class IV mutants induce IFN when overexpressed but do not

respond to c-di-GMP (Fig. 3a–c, blue, and Supplementary Table 1). Class V consists of mutants that have no effect on c-di-GMP binding or IFN induction (Fig. 3c, yellow, and Supplementary Table 1). Although mutating STING can result in diverse phenotypes, a key finding is that all mutants that failed to bind to c-di-GMP also lost the ability to induce IFN in response to c-di-GMP. Consistent with our observation that the CTD is sufficient for binding to c-di-GMP (Fig. 2f), all mutations that affected c-di-GMP binding were located within the CTD.

DNA and cyclic dinucleotides induce indistinguishable transcriptional responses in macrophages¹⁰, and STING seems to be essential for both responses^{4,11}. However, we found that STING expression is insufficient to restore the responsiveness of HEK293T cells to DNA, in contrast to cyclic dinucleotides (Fig. 1b and Supplementary Fig. 2a). Moreover, our competition assays indicate that DNA does not compete with cyclic-di-GMP for binding to STING under the conditions tested (Fig. 2d). Thus, although our data indicate that STING functions as a direct immunosensor of cyclic dinucleotides, it seems likely that additional host proteins are involved in IFN induction by DNA. Indeed, two candidate DNA sensors, DAI (also known as ZBP1) and IFI16, have been identified^{16,20}, neither of which seems to be essential for the response to cyclic dinucleotides (ref. 10 and data not shown). To determine whether the responsiveness to cyclic dinucleotides and DNA are separable functions of STING, we sought to identify STING mutants that fail to respond to cyclic dinucleotides but still respond to DNA. We identified a STING mutant (R231A) that was unresponsive to c-di-GMP (Fig. 4a), although it induced IFN when overexpressed (Fig. 4a) and bound to c-di-GMP (Fig. 4b). Interestingly, STING R231A was able to restore the responsiveness of *gt* bone marrow macrophages to DNA but not to cyclic-di-GMP (Fig. 4c). Thus, cyclic

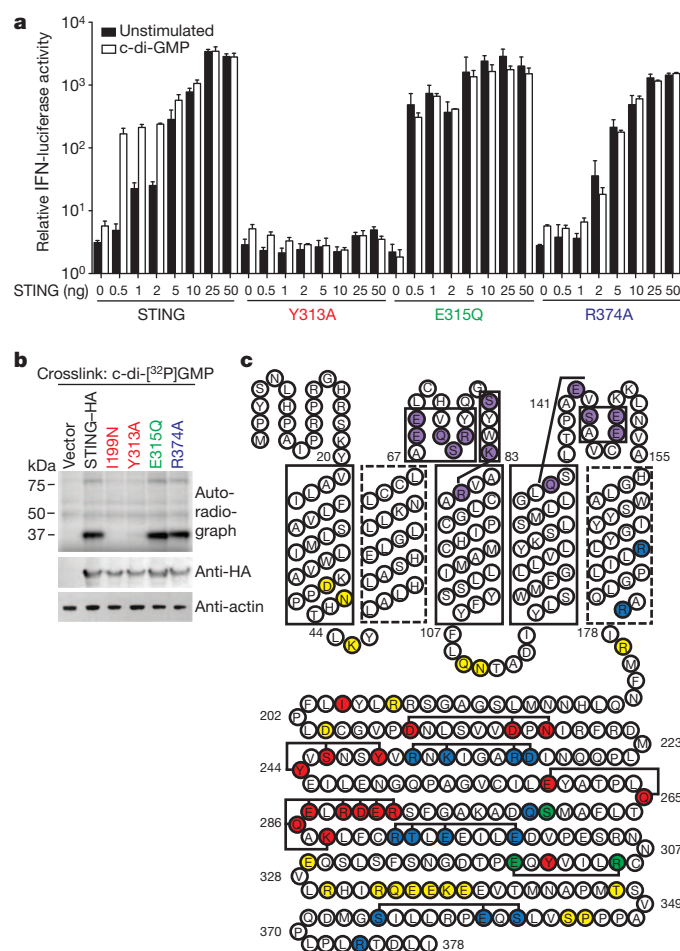


Figure 3 | Mutational analysis of STING. **a**, HEK293T cells were transfected with vectors encoding STING and STING mutants as indicated, together with an IFN-luciferase reporter, and luciferase activity was measured 6 h after stimulation. Data are presented as mean \pm s.d. ($n = 3$). **b**, HEK293T cells were transfected as in **a**, except that the lysates were subjected to the crosslinking assay with c-di-[³²P]GMP as in Fig. 2a. **c**, Organization of STING based on the membrane topology prediction programs SOSUI, TMHMM, HMMTOP and TMPred. Strongly predicted transmembrane domains are boxed; weakly predicted transmembrane domains have dashed boxes. Coloured residues indicate mutant classes (see text for description of mutants): class I (red), class II (purple), class III (green), class IV (blue) and class V (yellow). Bracketed mutations were made in combination. **a–c**, Data are representative of at least three independent experiments.

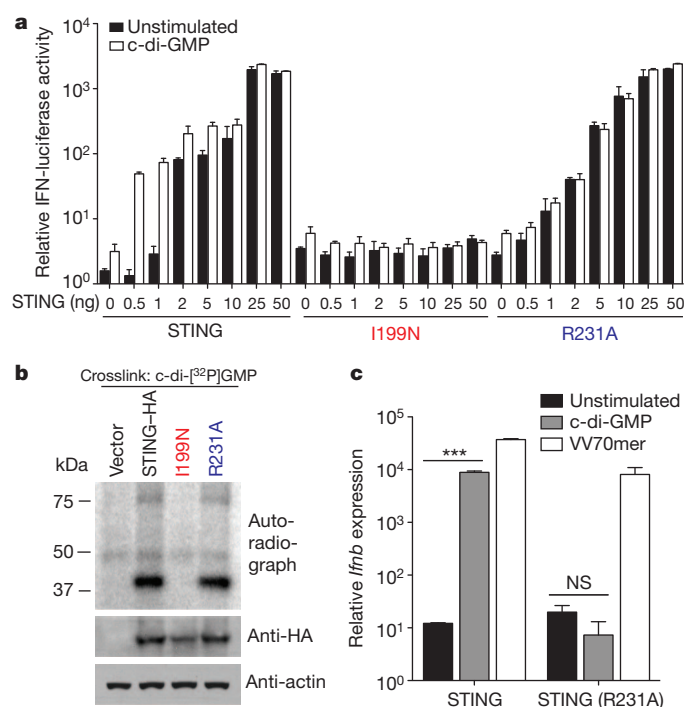


Figure 4 | The IFN response to DNA and c-di-GMP can be uncoupled. **a**, HEK293T cells were transfected as indicated, together with an IFN-luciferase reporter, and IFN reporter activity was measured 6 h after stimulation. **b**, HEK293T cells were transfected as in **a**, except that the lysates were subjected to the crosslinking assay with c-di-[³²P]GMP as in Fig. 2a. **c**, Bone-marrow-derived macrophages from *Sting*-deficient (*gt*) mice were transduced with the indicated constructs. IFN induction in response to transfected cyclic-di-GMP or VV70mer was measured by quantitative RT-PCR and normalized to *Rps17*. ***, $P < 0.001$; NS, not significant, $P = 0.1205$. **a, c**, Data are presented as mean \pm s.d. ($n = 3$). **a–c**, Data are representative of at least three independent experiments.

dinucleotide sensing and DNA sensing can be uncoupled, suggesting that these two pathways are discrete but share STING as a common signalling molecule. It is unexpected that STING would function both as a direct immunosensor (of cyclic dinucleotides) and as a signalling adaptor (in the response to DNA). One possibility is that STING initially evolved as a cyclic dinucleotide sensor and was subsequently co-opted for DNA sensing.

We previously used mouse mutagenesis to identify STING as an essential molecule in the *in vivo* IFN response to cyclic dinucleotides¹¹. The requirement for STING can now be rationalized by our proposal that STING functions as a direct sensor of cyclic dinucleotides. Interestingly, STING does not share homology with any known immunosensor and therefore seems to represent a novel category of microbial detector. Although a BLAST search of the mouse proteome for homologues of the *L. monocytogenes* diadenylate cyclase (Imo2120; also known as DacA) identifies STING as the top hit, the homology is limited to a short region of the STING CTD (amino acids 311–358). STING does not seem to share homology with PilZ-domain-containing proteins, which function as c-di-GMP receptors in bacteria⁸. Structural studies are required to better characterize the interaction of STING with cyclic dinucleotides and to determine whether STING resembles any known protein in mammals or bacteria.

Numerous studies have demonstrated that cyclic dinucleotides are potent immunostimulatory compounds that may be valuable as novel immunotherapeutics or adjuvants^{9,13}. The therapeutic development of cyclic dinucleotides will be greatly facilitated by an improved understanding of the mechanism by which they are sensed. Furthermore, our finding that STING is a direct detector of cyclic dinucleotides provides insight into the fundamental mechanisms by which the innate immune system can detect bacterial infection.

METHODS SUMMARY

Transfections. Transfections were carried out using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Digitonin permeabilization was used to introduce c-di-AMP into cells as described previously¹².

Recombinant STING. DNA encoding the CTD of mouse STING (nucleotides 414–1,137) was cloned into the vector pET28a for recombinant protein expression in *Escherichia coli*.

Ultraviolet radiation crosslinking. c-di-[³²P]GMP was enzymatically synthesized using recombinant WspR and was used in an ultraviolet radiation crosslinking assay as described previously²¹. Briefly, 50 µg HEK293T cell lysate at a final concentration of 2 µg µl⁻¹, or 1 µg recombinant His₆-tagged STING, was incubated with 2 µCi c-di-[³²P]GMP in binding buffer (20 mM Tris-HCl, pH 7.4, 200 mM NaCl and 1 mM MgCl₂) for 15 min at 25 °C. The reactions were irradiated at 254 nm, and the proteins in the samples were then separated by SDS-PAGE.

Statistical analysis. Statistical differences were calculated with an unpaired two-tailed Student's *t*-test using Prism 5.0b software (GraphPad).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 29 April; accepted 3 August 2011.

Published online 25 September 2011.

1. Ishii, K. J. *et al.* A Toll-like receptor-independent antiviral response induced by double-stranded B-form DNA. *Nature Immunol.* **7**, 40–48 (2006).
2. Stetson, D. B. & Medzhitov, R. Recognition of cytosolic DNA activates an IRF3-dependent innate immune response. *Immunity* **24**, 93–103 (2006).
3. Ishikawa, H. & Barber, G. N. STING is an endoplasmic reticulum adaptor that facilitates innate immune signalling. *Nature* **455**, 674–678 (2008).

4. Ishikawa, H., Ma, Z. & Barber, G. N. STING regulates intracellular DNA-mediated, type I interferon-dependent innate immunity. *Nature* **461**, 788–792 (2009).
5. Jin, L. *et al.* MPYS, a novel membrane tetraspanner, is associated with major histocompatibility complex class II and mediates transduction of apoptotic signals. *Mol. Cell. Biol.* **28**, 5014–5026 (2008).
6. Sun, W. *et al.* ERIS, an endoplasmic reticulum IFN stimulator, activates innate immune signaling through dimerization. *Proc. Natl Acad. Sci. USA* **106**, 8653–8658 (2009).
7. Zhong, B. *et al.* The adaptor protein MITA links virus-sensing receptors to IRF3 transcription factor activation. *Immunity* **29**, 538–550 (2008).
8. Tamayo, R., Pratt, J. T. & Camilli, A. Roles of cyclic diguanylate in the regulation of bacterial pathogenesis. *Annu. Rev. Microbiol.* **61**, 131–148 (2007).
9. Karaolis, D. K. *et al.* Bacterial c-di-GMP is an immunostimulatory molecule. *J. Immunol.* **178**, 2171–2181 (2007).
10. McWhirter, S. M. *et al.* A host type I interferon response is induced by cytosolic sensing of the bacterial second messenger cyclic-di-GMP. *J. Exp. Med.* **206**, 1899–1911 (2009).
11. Sauer, J. D. *et al.* The *N*-ethyl-*N*-nitrosourea-induced *Goldenticket* mouse mutant reveals an essential function of *Sting* in the *in vivo* interferon response to *Listeria monocytogenes* and cyclic dinucleotides. *Infect. Immun.* **79**, 688–694 (2011).
12. Woodward, J. J., Iavarone, A. T. & Portnoy, D. A. c-di-AMP secreted by intracellular *Listeria monocytogenes* activates a host type I interferon response. *Science* **328**, 1703–1705 (2010).
13. Chen, W., Kuolee, R. & Yan, H. The potential of 3',5'-cyclic diguanylic acid (c-di-GMP) as an effective vaccine adjuvant. *Vaccine* **28**, 3080–3085 (2010).
14. Witte, G., Hartung, S., Buttner, K. & Hopfner, K. P. Structural biochemistry of a bacterial checkpoint protein reveals diadenylate cyclase activity regulated by DNA recombination intermediates. *Mol. Cell* **30**, 167–178 (2008).
15. Janeway, C. A. Jr. Approaching the asymptote? Evolution and revolution in immunology. *Cold Spring Harb. Symp. Quant. Biol.* **54**, 1–13 (1989).
16. Unterholzner, L. *et al.* IFI16 is an innate immune sensor for intracellular DNA. *Nature Immunol.* **11**, 997–1004 (2010).
17. Ablasser, A. *et al.* RIG-I-dependent sensing of poly(dA:dT) through the induction of an RNA polymerase III-transcribed RNA intermediate. *Nature Immunol.* **10**, 1065–1072 (2009).
18. Chiu, Y. H., Macmillan, J. B. & Chen, Z. J. RNA polymerase III detects cytosolic DNA and induces type I interferons through the RIG-I pathway. *Cell* **138**, 576–591 (2009).
19. Chodosh, L. A. in *Current Protocols in Molecular Biology* Ch. 12, 12.5.1–12.5.8 (Wiley, 2001).
20. Takaoka, A. *et al.* DAI (DLM-1/ZBP1) is a cytosolic DNA sensor and an activator of innate immune response. *Nature* **448**, 501–505 (2007).
21. Merighi, M., Lee, V. T., Hyodo, M., Hayakawa, Y. & Lory, S. The second messenger bis-(3'-5')-cyclic-GMP and its PilZ domain-containing receptor Alg44 are required for alginate biosynthesis in *Pseudomonas aeruginosa*. *Mol. Microbiol.* **65**, 876–895 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank H. Carlson, K. Collins, S. McWhirter, D. Raulet, K. Sjölander and members of the Vance, Barton and Portnoy laboratories at the University of California, Berkeley, for advice and discussions. We thank J. Woodward and D. Portnoy for their gift of purified c-di-AMP. Work in R.E.V.'s laboratory is supported by investigator awards from the Burroughs Wellcome Fund and the Cancer Research Institute and by National Institutes of Health (NIH) grants AI075039, AI080749 and AI063302. D.L.B. is supported by an NIH National Research Service Award fellowship F32 (AI091100).

Author Contributions D.L.B. performed luciferase assays and quantitative RT-PCR, generated c-di-[³²P]GMP, purified recombinant STING, performed c-di-[³²P]GMP binding assays and transduced *gt* bone marrow macrophages. K.M.M. generated truncation mutations and performed luciferase assays. K.S.-T. generated point mutants and performed luciferase assays. D.L.B., K.M.M. and R.E.V. participated in study design and data analysis. D.L.B. and R.E.V. wrote the paper. B.E. contributed to protein purification methods. J.S.I. contributed to the design of the equilibrium dialysis experiments and the analysis of binding data. M.H. and Y.H. synthesized c-di-GMP.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.E.V. (rvance@berkeley.edu).

METHODS

Synthesis of c-di- 32 P]GMP. The synthesis of c-di- 32 P]GMP was carried out as described previously²¹. Briefly, recombinant His₆-tagged WspR was incubated with [α - 32 P]GTP (3,000 Ci mmol⁻¹, 10 μ Ci μ l⁻¹, Amersham Biosciences) for 2 h at 25 °C, followed by heat inactivation of WspR at 95 °C for 5 min. Residual 32 P was removed by incubation with calf intestinal phosphatase (CIP) (New England Biolabs) for 10 min at 37 °C. CIP was heat inactivated at 95 °C for 5 min, followed by centrifugation at 16,000g for 5 min. The [32 P]GTP used as a negative control was prepared identically except that His₆-tagged WspR and CIP were omitted from the preparation. Radiolabelled nucleotides were quantified by separation by thin-layer chromatography on cellulose-PEI plates (Macherey-Nagel) using 1.5 M KH₂PO₄, pH 3.65 (Supplementary Fig. 5).

Cell lines and animals. C57BL/6 *Myd88*^{-/-} *Trif*^{-/-} mice were obtained from G. Barton, and immortalized macrophages were generated as described previously²². Immortalized bone marrow macrophages were maintained in RPMI 1640 (Invitrogen) supplemented with 10% FBS, penicillin-streptomycin and glutamine. HEK293T cells were maintained in DMEM supplemented with 10% FBS, penicillin-streptomycin and glutamine. Animal use was approved by the Animal Care and Use Committee at the University of California, Berkeley.

Plasmids. A construct encoding RocR (NP_252636) from *P. aeruginosa*²³ was a gift from S. Lory. *rocR* was cloned into the MSCV2.2 retroviral expression construct upstream of an internal ribosome entry site (IRES)-green fluorescent protein (GFP). MSCV-*rocR* was transduced into immortalized macrophages from *Myd88*^{-/-} *Trif*^{-/-} mice, and cells were sorted for GFP expression. Mouse *Sting* and the *gt* (I199N) mutant allele of *Sting* were cloned into the vector pcDNA3 with a C-terminal HA tag as described previously¹¹. DNA encoding the CTD of mouse *Sting* (nucleotides 414–1,137) was cloned into pET28a for recombinant protein expression in *Escherichia coli*.

Site-directed mutagenesis. Mutations in *Sting* were generated using the QuikChange Site-Directed Mutagenesis Kit (Stratagene) according to the manufacturer's guidelines.

Reagents. c-di-GMP was synthesized as described previously²⁴. Purified c-di-AMP was a gift from J. Woodward and D. Portnoy. Poly(dAT:dTA), GTP, ATP, GMP and guanosine were obtained from Sigma-Aldrich. Poly(I)•poly(C) (denoted poly(I:C)) was purchased from Invivogen. Guanosine-3',5'-bisdiphosphate (ppGpp) was obtained from TriLink Biotechnologies. Sendai virus was purchased from Charles River Laboratories. Theiler's murine encephalomyelitis virus (TMEV) strain GDVII was provided by M. Brahic and E. Freundt. DNA oligonucleotides corresponding to the VV70mer and ISD were purchased from Elim Biopharmaceuticals and were annealed as described previously^{2,16}.

Cell stimulation. All transfections (excluding for c-di-AMP) were carried out using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. The VV70mer was transfected at a final concentration of 0.5 μ g ml⁻¹. Poly(dAT:dTA), poly(I:C) and c-di-GMP were transfected at a final concentration of 4 μ g ml⁻¹. c-di-AMP was used at a final concentration of 5.4 mM, and stimulation was performed using digitonin permeabilization as described previously¹².

Luciferase assay. HEK293T cells were plated in TC-treated 96-well plates at 0.5×10^6 cells ml⁻¹. The next day, the cells were transfected as indicated, together with IFN- β -firefly luciferase and TK-*Renilla* luciferase reporter constructs. Following stimulation for 6 h with the indicated ligands, the cells were lysed in passive lysis buffer (Promega) for 5 min at 25 °C. The cell lysates were incubated with firefly luciferase substrate (Biosynth) and the *Renilla* luciferase substrate coelenterazine (Biotium), and luminescence was measured on a SpectraMax L microplate reader (Molecular Devices). The relative *Ifnb* expression was calculated as firefly luminescence relative to *Renilla* luminescence.

Quantitative PCR. The analysis of *Ifnb* expression by bone marrow macrophages was conducted as described previously¹⁰.

Preparation of HEK293T cell lysates and immunoprecipitations. HEK293T cells were plated at a density of 1×10^6 cells well⁻¹ in a 6-well plate. The following day, the cells were transfected with pcDNA3 or pcDNA3 expressing HA-tagged

wild-type or mutant STING using Lipofectamine 2000 (Invitrogen). The day after that, the cells were rinsed once with PBS and transferred to Eppendorf tubes in PBS containing 1 mM EDTA. The cells were pelleted briefly by centrifugation at 1,000g at 4 °C. The cell pellet was lysed in an equal volume of digitonin lysis buffer (0.5% digitonin, 20 mM Tris-HCl, pH 7.4, and 150 mM NaCl) containing protease inhibitors (Roche) for 10 min on ice. The cell lysates were centrifuged at 10,000g for 10 min at 4 °C. The protein concentration in the resultant supernatant was measured using the Bradford reagent (Bio-Rad). The cell lysates were subjected to a c-di-GMP binding (crosslinking) assay (see below). The lysates were separated by SDS-PAGE, and the separated proteins were transferred to a nitrocellulose membrane, which was then probed with rat anti-HA antibodies (Roche), to confirm STING-HA expression, and mouse anti- β -actin antibodies (Santa Cruz Biotechnology). To immunoprecipitate HA-tagged STING, the cell lysates were prepared similarly in digitonin lysis buffer and incubated with anti-HA-antibody-conjugated agarose beads (Sigma) for 2 h at 4 °C. Washed beads were subjected to a c-di-GMP binding assay or separated by SDS-PAGE and stained with colloidal blue protein stain (Thermo Scientific).

c-di-GMP binding assays. The c-di-GMP binding assay (also called the crosslinking assay) was based on a method described previously²¹. Briefly, 50 μ g HEK293T cell lysate at a final concentration of 2 μ g μ l⁻¹, or 1 μ g recombinant His₆-tagged STING, was incubated with 2 μ Ci radiolabelled nucleotide in binding buffer (20 mM Tris-HCl, pH 7.4, 200 mM NaCl and 1 mM MgCl₂) for 15 min at 25 °C. The reactions were irradiated at 254 nm for 20 min on ice at a 3-cm distance from a UVG-54 mineral light lamp (UVP). Immediately after crosslinking, the reactions were terminated by the addition of SDS sample buffer (40% glycerol, 8% SDS, 2% 2-mercaptoethanol, 40 mM EDTA, 0.05% bromophenol blue and 250 mM Tris-HCl, pH 6.8), boiled for 5 min and then separated by SDS-PAGE. The gels were dried, exposed to a phosphor screen and visualized using a Typhoon Trio imager (GE Healthcare).

Protein purification. The construct expressing a constitutively active form of WspR (pQE-WspR*)^{23,25} was a gift from S. Lory. The purification of His₆-tagged WspR was carried out as described previously, using Ni-NTA affinity chromatography (Qiagen)²⁶. DNA encoding the CTD of mouse STING (nucleotides 414–1,137) was cloned into the vector pET28a and purified by Ni-NTA affinity chromatography (Qiagen) according to the manufacturer's instructions.

Equilibrium dialysis. The binding affinity of radioactive c-di-GMP was measured by equilibrium dialysis, using a 96-well equilibrium dialyser (Harvard Apparatus) with a 5,000 molecular weight cut-off membrane. One chamber contained 150 μ l 10 μ M purified His₆-tagged STING(138–378) in assay buffer (25 mM Tris-HCl, pH 7.4, 100 mM NaCl, 1 mM MgCl₂ and 10% glycerol), while the other was filled with 150 μ l c-di- 32 P]GMP at a range of concentrations (40–160 μ M). Equilibrium was reached after 48 h at 25 °C, and three samples were drawn from each chamber and mixed with 2 ml Econo-Safe scintillation fluid. Samples were measured in an LS 6000 IC scintillation counter (Beckman). Data analysis was performed using Prism 5.0b software (GraphPad). The dissociation constant (K_d), the maximum number of binding sites (B_{max}) and the Hill coefficient (h) were generated using nonlinear regression, allowing one-site specific binding with a Hill slope.

22. Broz, P., von Moltke, J., Jones, J. W., Vance, R. E. & Monack, D. M. Differential requirement for caspase-1 autoproteolysis in pathogen-induced cell death and cytokine processing. *Cell Host Microbe* **8**, 471–483 (2010).
23. Kulesekara, H. et al. Analysis of *Pseudomonas aeruginosa* diguanylate cyclases and phosphodiesterases reveals a role for bis-(3'-5')-cyclic-GMP in virulence. *Proc. Natl Acad. Sci. USA* **103**, 2839–2844 (2006).
24. Hyodo, M. & Hayakawa, Y. An improved method for synthesizing cyclic bis(3'-5')diguanylic acid (c-di-GMP). *Bull. Chem. Soc. Jpn* **77**, 2089–2093 (2004).
25. Simm, R., Remminghorst, U., Ahmad, I., Zakikhany, K. & Romling, U. A role for the EAL-like protein STM1344 in regulation of CsgD expression and motility in *Salmonella enterica* serovar Typhimurium. *J. Bacteriol.* **191**, 3928–3937 (2009).
26. De, N. et al. Phosphorylation-independent regulation of the diguanylate cyclase WspR. *PLoS Biol.* **6**, e67 (2008).

Control of flowering and storage organ formation in potato by FLOWERING LOCUS T

Cristina Navarro^{1*}, José A. Abelenda^{1*}, Eduard Cruz-Oró¹, Carlos A. Cuéllar¹, Shojiro Tamaki², Javier Silva¹, Ko Shimamoto² & Salomé Prat¹

Seasonal fluctuations in day length regulate important aspects of plant development such as the flowering transition or, in potato (*Solanum tuberosum*), the formation of tubers. Day length is sensed by the leaves, which produce a mobile signal transported to the shoot apex or underground stems to induce a flowering transition or, respectively, a tuberization transition. Work in *Arabidopsis*, tomato and rice (*Oryza sativa*) identified the mobile FLOWERING LOCUS T (FT) protein as a main component of the long-range 'florigen', or flowering hormone, signal^{1–3}. Here we show that expression of the *Hd3a* gene, the FT orthologue in rice, induces strict short-day potato types⁴ to tuberize in long days. Tuber induction is graft transmissible and the *Hd3a*–GFP protein is detected in the stolons of grafted plants, transport of the fusion protein thus correlating with tuber formation. We provide evidence showing that the potato floral and tuberization transitions are controlled by two different FT-like paralogues (StSP3D and StSP6A) that respond to independent environmental cues, and show that an autorelay mechanism involving CONSTANS modulates expression of the tuberization-control StSP6A gene.

Potato, the third largest global food crop after wheat and rice, is cultivated for its underground storage stems or tubers, which are rich in starch and other nutrients. Short days and cool temperatures promote tuber formation, ensuring that differentiation of these vegetative propagation organs precedes winter. Whereas cultivated potatoes derive from Chilean landraces more adapted to long-day conditions, Andean types such as the *S. tuberosum* group Andigena (*Solanum tuberosum andigena*) tuberize only in short days⁵. These plants require night periods longer than a critical length, as a pulse of light during the night (a 'night break') represses tuberization, as seen in strict short-day flowering plants^{6,7}.

Day length is sensed by expanded leaves, which synthesize a mobile signal or 'tuberigen' that is transported to the underground stems to induce tuber formation⁴. This long-distance signal shares several features with the mobile florigen, with different pieces of evidence suggesting that related photoperiodic pathways may control synthesis of both signals⁸. In *Arabidopsis thaliana*, activation of the FT gene by CONSTANS (CO) mediates floral transition in long days^{9,10}. Transport of the FT protein from the leaf to the shoot apical meristem has been demonstrated in diverse ways^{11–13}, it being broadly accepted that this phloem mobile protein functions as the florigen. Closely related genes also mediate control of flowering in rice, although in this short-day plant the CO orthologue *Heading date 1*¹⁴ (*Hd1*) activates expression of the FT-like *Hd3a* gene in short days but represses its expression under long-day conditions^{15,16}.

A CO-dependent pathway is also thought to mediate short-day tuberization, as expression of *Arabidopsis* CO in Andigena plants delays tuber formation in short days¹⁷. High light irradiance, however, induces Andigena potatoes to flower in long days, although these conditions are restrictive for tuberization⁸. This long-day flowering

response seemed to argue against a role for FT in tuberization, implicating an additional CO target as the mobile tuberigen. Here we show that ectopic expression of the rice *Hd3a* gene induces Andigena plants to flower and tuberize under non-inductive long days, demonstrating the potential of FT to act as the mobile tuberigen. We show, in addition, that flowering and short-day tuberization responses are regulated by two members of the potato FT-like gene family that respond to different environmental cues.

To assess whether FT has a role in tuberization, we transformed Andigena plants with the *rolC::Hd3a*–GFP construct (GFP, green fluorescent protein), which in rice promotes floral transition in long days¹⁸. Lines expressing this construct were induced to flower (Fig. 1b, c) and were able to tuberize in non-inductive long days (Fig. 1a). When grafted to wild-type plants, these lines induced the wild-type controls to tuberize in long days, independently of whether they were used as donors (*Hd3a* onto wild type) or as stocks (wild type onto *Hd3a*), whereas none of the control grafts (wild type onto wild type; Fig. 1d) tuberized in non-inductive long days. We detected the *Hd3a*–GFP protein but not its transcript in the stolons of grafted wild-type stocks (Supplementary Fig. 1a, c), demonstrating that the protein but not the RNA can move across the graft junction and function as a powerful tuberization inducer.

Studies in tomato identified six members of the *SELF-PRUNING* (SP) gene family¹⁹ and sequencing of two diploid potato (Tuberosum RH89-039-16 and Phureja DM1-3 516 R44) genomes recently led to the identification of three additional FT and TFL family members²⁰. Phylogeny of these genes grouped the StSP6A, StSP5G, StSP5G-like and StSP3D homologues into the same clade as the *Arabidopsis* FT, tomato *SINGLE-FLOWER TRUSS* (SFT) and rice *Hd3a* genes (Supplementary Fig. 2). Quantitative PCR with reverse transcription (RT–PCR) revealed that these transcripts are expressed in leaves or stolons, whereas transcripts for *SP/TFL1* and *MFT* members are more ubiquitously distributed (Fig. 1e). Interestingly, StSP6A gene expression strongly correlates with tuberization, high levels of expression being observed in leaves and stolons of short-day-induced plants and in antisense phytochrome B lines, with constitutive tuberization²¹ (Fig. 1f and Supplementary Information, section 1). This expression profile suggests that this gene is involved in tuberization control, StSP6A overexpression (StSP6Aox) lines being able to tuberize under non-inductive long days (Fig. 2a, b) and induced to flower, although their flowering phenotype is less severe than in *Hd3a* lines (Supplementary Fig. 3a, b and Supplementary Information, section 2). StSP6A silencing, in turn, strongly delays tuber formation in short days (Fig. 2c, d), pointing to an essential role for this FT-like protein in tuberization promotion. StSP6A expression analyses in commercial cultivars with early (Jaerla), late (Baraka) and intermediate (Kennebec) tuberization periods, in addition, show that accumulation of this transcript in leaves correlates with the tuberization time of these cultivars (Supplementary Fig. 6), indicating that this FT paralogue is involved in tuberization control even in non-photoperiodic cultivars.

¹Departamento de Genética Molecular de Plantas, Centro Nacional de Biotecnología-CSIC. Calle Darwin 3, 28049 Madrid, Spain. ²Laboratory of Plant Molecular Genetics, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma 630-0101, Japan.

*These authors contributed equally to this work.

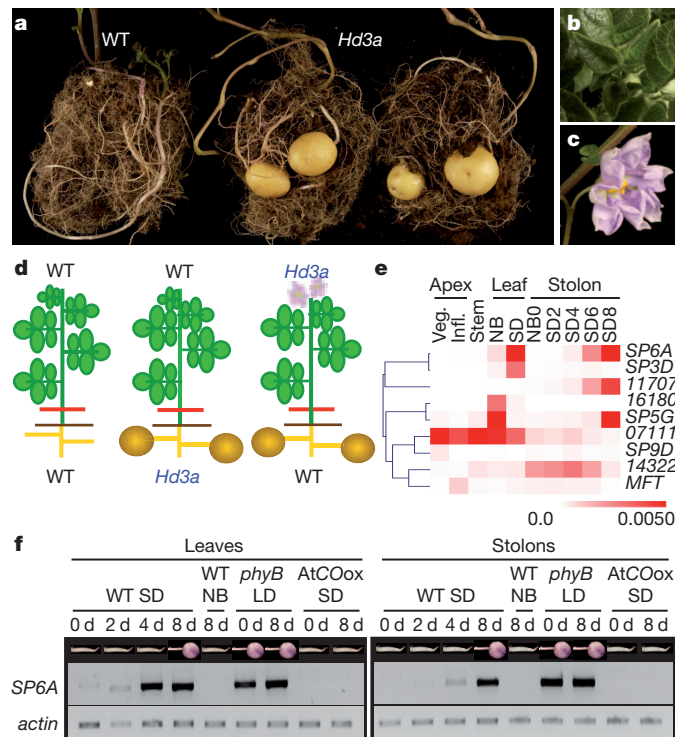


Figure 1 | Phenotype of *Andigena rolC::Hd3a*-GFP lines and expression profiles of the potato *FT*- and *TFL1*-like genes. **a, *Hd3a* lines (centre and right) tuberize under long-day conditions. Left, wild-type (WT) control. **b**, **c**, Early-flowering *Hd3a* plants (**c**) relative to wild type (**b**). **d**, Tuber induction in long days of wild-type plants grafted with *Hd3a* donors (right) or stocks (centre). Wild-type control grafts do not tuberize (left). Red lines indicate the graft junction. **e**, Relative levels of expression of the potato *FT*- and *TFL1*-like genes. 11707, PGSC0003DMG400011707; 16180, PGSC0003DMG400016180; 07111, PGSC0003DMG400007111; 14322, PGSC0003DMG400014322; Veg., vegetative apex; Inf., inflorescence apex; NB, night break; SD, short day. **f**, Semiquantitative RT-PCR analysis of *StSP6A* expression in leaves and stolons of plants with different tuberization states (indicated on top). The pictures show non-induced stolons or tubers. LD, long day.**

Day-neutral tomato flowering is regulated by the *FT*-homologue *SFT/SP3D* gene². This gene has been reported to be regulated independently of CO and day length^{22,23}, its potato orthologue being likely to have a role in light-irradiance-dependent flowering. *StSP3D*-silenced lines were actually found to be late flowering (Fig. 2e–h), although under short-day conditions they tuberized at the same time as untransformed controls (Supplementary Fig. 7 and Supplementary Information, section 2), indicating that this gene is important in flowering but not in the tuberization transition. Remarkably, in modern tomato cultivars *SP6A* was reported to have an extra nucleotide (T at position 421) that leads to a premature stop codon, implying a non-functional role for this gene¹⁹. *StSP3D*, in turn, retains a weak short-day activation response in *Andigena*, although transcript levels are lower than those of *StSP6A* (Fig. 1e). Together, these observations suggest that members of the *FT*-like family in Solanaceae diversified such that *SFT/SP3D* ceased to be regulated by CO^{22,23} and became responsive to other environmental cues, such as to be important in day-neutral flowering control. *SP6A* function was lost in modern tomato cultivars but in potato plays a major part in tuberization, as day-length-dependent activation of this gene mediates strict short-day tuberization of *Andigena* species. Thus, the roles of these two *FT* paralogues in flowering and tuberization control have evolved, at least in part, through changes in their expression profiles, both genes encoding for functionally similar proteins, as indicated by the finding that *StSP6A* expression in *Arabidopsis* rescues the late-flowering phenotype of *co-1* and *ft-1* mutants (Supplementary Fig. 8).

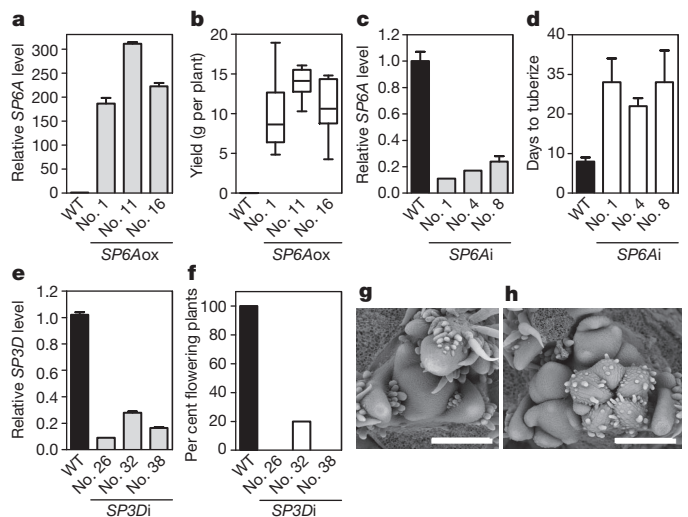


Figure 2 | Function of the *StSP6A* and *StSP3D* genes in short-day-dependent tuberization and day-neutral flowering control. **a, **b**, *StSP6A* expression in leaves of *StSP6Aox* lines (**a**) and tuber induction in these lines (**b**) in long days. **c**, **d**, *StSP6A* silencing in leaves of *StSP6A* RNAi (*SP6Ai*) lines (**c**) correlates with late tuberization in short days (**d**). **e**, Relative *StSP3D* expression in *StSP3D* RNAi (*SP3Di*) lines. Error bars, s.d. ($n > 3$). **f**, Late flowering of *StSP3D*-silenced lines. **g**, **h**, Scanning electron microscopy images of the *StSP3D* RNAi apices: vegetative apex in L26 (**g**); apex in inflorescence transition in L32 (**h**). Scale bars, 200 μ m.**

Interestingly, two other *FT*-clade members, *StSP5G* and *StSP5G*-like (Supplementary Table 2), are expressed under non-inductive long-day conditions (Fig. 1e). This expression profile suggests an antagonistic function for *StSP6A* and these two *FT* paralogues (Supplementary Information, section 3), as recently reported for the sugar beet (*Beta vulgaris*) *BvFT1* and *BvFT2* genes²⁴. Further studies will be required to assess whether these *FT*-clade members have an inhibitory effect on tuberization.

StSP6A is expressed not only in leaves but also in stolons of tuberizing plants (Fig. 1f). Expression in these organs is delayed with respect to the leaves, suggesting an autoregulatory loop for the transported protein. In support of this relay mechanism, increased levels of expression of the endogenous *StSP6A* transcript are observed in *Hd3a* lines (Supplementary Fig. 10) and wild-type stocks grafted to these plants (Supplementary Fig. 10a). Thus, in contrast to the *Arabidopsis* *FT* or rice *Hd3a* gene, *StSP6A* is regulated by a relay mechanism that sustains synthesis of the inducing signal in stolons. In this regard, a local balance between the SP floral repressor and the tomato *SFT* florigen signal has been shown to contribute to the differential flowering response of primary and secondary shoot meristems²⁵. Hence, it is possible that having more SP in stolons confers a reduced sensitivity to FT, which may explain why floral transition is activated without tuber formation, whereas *Hd3a* overexpression activates both developmental processes.

To rule out that tuberization of grafted stock plants is mostly mediated by this relay mechanism, we grafted *Hd3a* grafts onto *StSP6A* RNA interference (RNAi) stocks to test whether inhibition of *StSP6A* expression blocks *Hd3a* signalling. As shown in Fig. 3a, activation of the endogenous *StSP6A* gene is strongly reduced in *StSP6A* RNAi compared with the wild-type stocks. Tuberization onset is delayed and tuber yield reduced in RNAi lines relative to the *Hd3a*/wild-type grafted controls (Fig. 3b), owing to impaired signal amplification, but this does not preclude tuberization of the RNAi stocks, highlighting induction by the *Hd3a* protein.

Finally, we tested whether this regulatory relay requires CO function, by grafting *StSP6Aox* plants into *StCOox* stocks and analysing endogenous *StSP6A* gene expression (Fig. 3c–e). Like the rice Hd1 protein^{15,16}, *StCO* represses *StSP6A* gene expression in long days, and

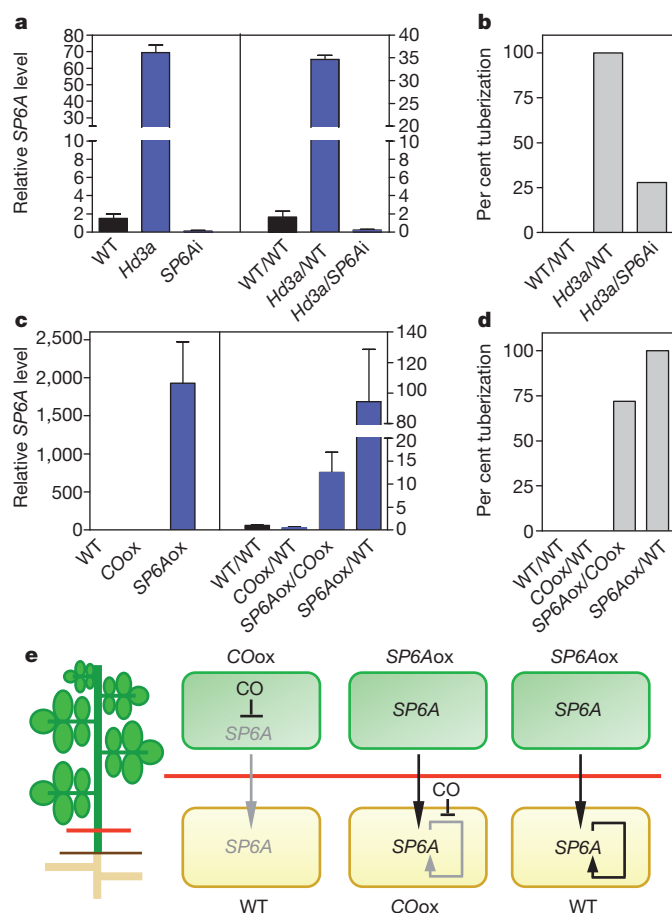


Figure 3 | Autoregulatory loop for StSP6A expression. **a**, Levels of StSP6A transcript in transgenic (left) and grafted (right) plants. *Hd3a* scions were grafted onto wild-type or StSP6A RNAi (SP6Ai) lines. **b**, Tuber induction in grafted plants, scored when 100% of the *Hd3a*/wild-type grafts were tuberizing. **c**, Relative expression of the endogenous StSP6A transcript in COox and SP6Aox plants (left), and in grafts with wild-type and COox stocks (right). **d**, Tuber induction of the grafted plants grown in non-inductive long days. Error bars, s.d. of two experimental replicates ($n > 8$). See Supplementary Fig. 12 for each replicate result. **e**, StSP6A autoregulation.

transfer to short-day conditions relieves this repression (Supplementary Information, section 4). In line with this function, activation of the StSP6A gene is largely repressed in stolons of the StSP6Aox/StCOox grafted plants (Fig. 3c), relative to StSP6Aox/wild-type grafts used as controls, which implies that StCO is involved in the autoregulatory loop that drives StSP6A expression in stolons. Moreover, StSP6A accumulates only to basal levels in the stolons of StCOox/wild-type and wild-type/wild-type grafts used as negative controls (StSP6A is not expressed in long days in StCOox or wild-type scions), which implies that expression in these organs requires the mobile protein produced in the leaves.

Our data provide the molecular basis for a long-standing physiological observation, namely that flowering tobacco plants grafted onto non-induced potato stocks induce tuberization of the stock plants, irrespective of the photoperiodic requirement of the donor plant²⁶. An additional issue is how flowering and tuberization transitions are differentially triggered in response to the mobile FT signal. In *Arabidopsis*, FT interacts with FD to activate expression of floral meristem identity genes^{27,28}. We observed that *Hd3a* stolons initiate floral buds from the apical meristem (Fig. 4a–c) at the same time as they differentiate tubers from the subapical region. Using transgenic lines expressing the StSP6A protein under control of an ethanol-inducible promoter, tuber-specific transcripts were observed within 4 h of induction in the stolons (Fig. 4d, e and Supplementary Information,

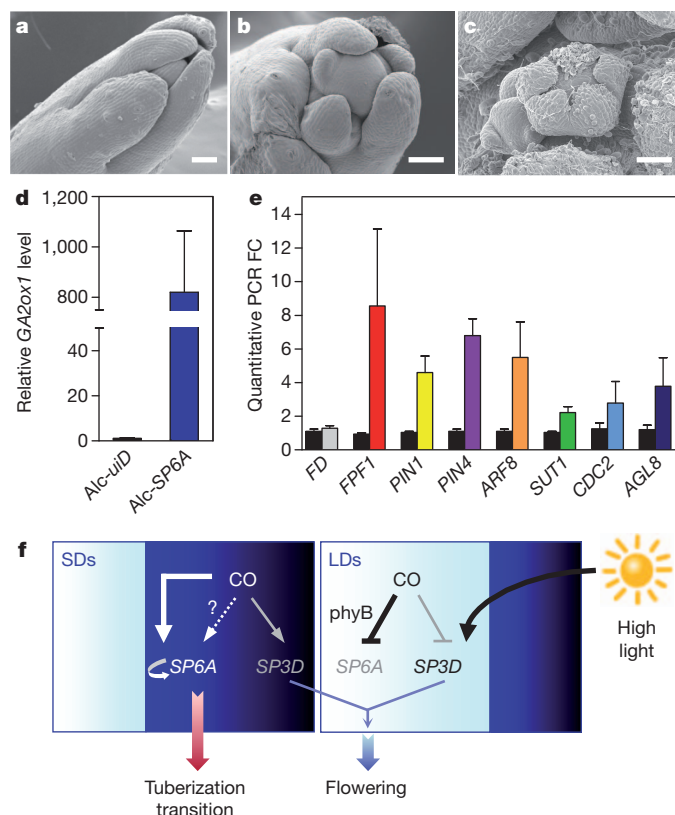


Figure 4 | Floral phenotype of *Hd3a* stolons and proposed model for flower and tuber induction. **a–c**, Flower induction in *Hd3a* stolons. Scanning electron microscope images of wild-type (**a**) and *Hd3a* (**b**, **c**) stolon apical meristems. Scale bars, 100 μ m. **d**, Local StSP6A induction activates StGA2ox1 gene expression in the stolon. **e**, Activation of other genes reported to be induced early during tuber development. Colour bars represent fold change (FC) in gene expression in Alc-StSP6A stolons relative to Alc-uid controls (black bars). Alc, ethanol-inducible promoter; error bars, s.d. of two experimental replicates ($n > 10$). **f**, Model for regulation of flowering and tuberization transitions. See Supplementary Information, section 6, for a detailed explanation.

section 5). This very rapid response precludes transport of any mobile signal from the leaves and thus supports the notion that the StSP6A protein is the mobile tuberigen transported to below-ground organs. Thus, it is possible that StSP6A interacts in stolon subapical cells with an as yet unknown transcriptional regulator, or that FD–StSP6A differentially activates a set of target genes specific to these cells, such as StGA2ox1 (ref. 29; Fig. 4d). A challenge for the future is to identify such stolon-specific StSP6A target genes and/or StSP6A-interacting partners, to establish how formation of these storage organs is regulated.

Identification of FT as a switch for tuberization supports the notion that FT function extends beyond flowering induction (Fig. 4f). Other studies have implicated FT in seasonal control of growth cessation in poplar trees³⁰ and meristem growth termination in tomato²⁵. Thus, FT is emerging as a key mobile signal controlling not only flowering but also a number of other meristem-associated transitions.

METHODS SUMMARY

Plant materials and growth conditions. Andigena 7540 wild-type plants and antisense phytochrome B²¹ and AtCOox (ref. 17) lines were grown in a greenhouse under long-day conditions. Growth chambers were used for controlled short-day (8 h light, 16 h dark) and short-day/night-break (short day plus a 30-min pulse of light in the middle of the night) treatments. Transgenic Andigena plants were generated by *Agrobacterium*-mediated transformation of leaf explants. Plants were grafted as previously described¹⁷ and cultivated under long-day conditions to analyse their tuberization response.

Plasmid constructs. StSP6A was cloned into the pBinAR binary vector to generate StSP6A-overexpressing lines. StSP6A, StSP3D and StCO RNAi constructs were

generated by inserting the 3' non-conserved regions of these genes in opposite orientations into the pBIN19RNAi vector. The pGWB2 vector was used for StCO overexpression. StSP6A ethanol-inducible lines were generated by transformation with the pBinSRNA-GW plasmid containing the StSP6A coding region. Primer sets used for these constructs are listed in Supplementary Table 3.

Real-time RT-PCR analyses. First-strand complementary DNA was synthesized from 2 µg total RNA and 1 µl of the reaction used for real-time gene expression analysis with the SYBR Green PCR master mix (Applied Biosystems). The *actin 8* gene was used for normalization. Identical procedures were used for semiquantitative RT-PCR, except that amplification was conducted in a Peltier thermal cycler (PTC-200, MJ Research). For sets of specific primers and product lengths, see Supplementary Table 3.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 9 December 2010; accepted 8 August 2011.

Published online 25 September 2011.

- Turck, F., Fornara, F. & Coupland, G. Regulation and identity of florigen: FLOWERING LOCUS T moves center stage. *Annu. Rev. Plant Biol.* **59**, 573–594 (2008).
- Lifschitz, E. *et al.* The tomato FT ortholog triggers systemic signals that regulate growth and flowering and substitute for diverse environmental stimuli. *Proc. Natl Acad. Sci. USA* **103**, 6398–6403 (2006).
- Zeevaert, J. A. Leaf-produced floral signals. *Curr. Opin. Plant Biol.* **11**, 541–547 (2008).
- Jackson, S. D. Multiple signalling pathways control tuber induction in potato. *Plant Physiol.* **119**, 1–8 (1999).
- Ewing, E. E. & Struik, P. C. Tuber formation in potato: induction, initiation and growth. *Hortic. Rev. (Am. Soc. Hortic. Sci.)* **14**, 89–197 (1992).
- Ishikawa, R. *et al.* Suppression of the floral activator Hd3a is the principal cause of the night break effect in rice. *Plant Cell* **17**, 3326–3336 (2005).
- Lagercrantz, U. At the end of the day: a common molecular mechanism for photoperiod responses in plants? *J. Exp. Bot.* **60**, 2501–2515 (2009).
- Rodríguez-Falcón, M., Bou, J. & Prat, S. Seasonal control of tuberization in potato: conserved elements with the flowering response. *Annu. Rev. Plant Biol.* **57**, 151–180 (2006).
- Samach, A. *et al.* Distinct roles of CONSTANS target genes in reproductive development of *Arabidopsis*. *Science* **288**, 1613–1616 (2000).
- An, H. *et al.* CONSTANS acts in the phloem to regulate a systemic signal that induces photoperiodic flowering of *Arabidopsis*. *Development* **131**, 3615–3626 (2004).
- Corbesier, L. *et al.* FT protein movement contributes to long-distance signaling in floral induction of *Arabidopsis*. *Science* **316**, 1030–1033 (2007).
- Jaeger, K. E. & Wigge, P. A. FT protein acts as a long-range signal in *Arabidopsis*. *Curr. Biol.* **17**, 1050–1054 (2007).
- Mathieu, J., Warthmann, N., Kuttner, F. & Schmid, M. Export of FT protein from phloem companion cells is sufficient for floral induction in *Arabidopsis*. *Curr. Biol.* **17**, 1055–1060 (2007).
- Yano, M. *et al.* Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the *Arabidopsis* flowering time gene CONSTANS. *Plant Cell* **12**, 2473–2484 (2000).
- Kojima, S. *et al.* Hd3a, a rice ortholog of the *Arabidopsis* FT gene, promotes transition to flowering downstream of Hd1 under short-day conditions. *Plant Cell Physiol.* **43**, 1096–1105 (2002).
- Hayama, R., Yokoi, S., Tamaki, S., Yano, M. & Shimamoto, K. Adaptation of photoperiodic control pathways produces short-day flowering in rice. *Nature* **422**, 719–722 (2003).
- Martinez-Garcia, J. F., Virgos-Soler, A. & Prat, S. Control of photoperiod-regulated tuberization in potato by the *Arabidopsis* flowering-time gene CONSTANS. *Proc. Natl Acad. Sci. USA* **99**, 15211–15216 (2002).
- Tamaki, S., Matsuo, S., Wong, H. L., Yokoi, S. & Shimamoto, K. Hd3a protein is a mobile flowering signal in rice. *Science* **316**, 1033–1036 (2007).
- Carmel-Goren, L., Liu, Y. S., Lifschitz, E. & Zamir, D. The SELF-PRUNING gene family in tomato. *Plant Mol. Biol.* **52**, 1215–1222 (2003).
- The Potato Genome Sequencing Consortium. Genome sequence and analysis of the tuber crop potato. *Nature* **475**, 189–195 (2011).
- Jackson, S. D., Heyer, A., Dietze, J. & Prat, S. Phytochrome B mediates the photoperiodic control of tuber formation in potato. *Plant J.* **9**, 159–166 (1996).
- Ben-Naim, O. *et al.* The CCAAT binding factor can mediate interactions between CONSTANS-like proteins and DNA. *Plant J.* **46**, 462–476 (2006).
- Lifschitz, E. & Eshed, Y. Universal florigenic signals triggered by FT homologues regulate growth and flowering cycles in perennial day-neutral tomato. *J. Exp. Bot.* **57**, 3405–3414 (2006).
- Pin, P. A. *et al.* An antagonistic pair of FT homologs mediates the control of flowering time in sugar beet. *Science* **330**, 1397–1400 (2010).
- Shalit, A. *et al.* The flowering hormone florigen functions as a general systemic regulator of growth and termination. *Proc. Natl Acad. Sci. USA* **106**, 8392–8397 (2009).
- Chailakhyan, M. K., Yanina, L. I., Davedzhyan, A. G. & Lotova, G. N. Photoperiodism and tuber formation in grafting of tobacco onto potato. *Dokl. Akad. Nauk SSSR* **257**, 1276–1280 (1981).
- Abe, M. *et al.* FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. *Science* **309**, 1052–1056 (2005).
- Wigge, P. A. *et al.* Integration of spatial and temporal information during floral induction in *Arabidopsis*. *Science* **309**, 1056–1059 (2005).
- Kloosterman, B. *et al.* StGA2ox1 is induced prior to stolon swelling and controls GA levels during potato tuber development. *Plant J.* **52**, 362–373 (2007).
- Bohlenius, H. *et al.* CO/FT regulatory module controls timing of flowering and seasonal growth cessation in trees. *Science* **312**, 1040–1043 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Paz-Ares, G. Bryan and C. Bachem for their comments on the manuscript. We also thank S. Yokoi for his help in this work. This research was supported by grants from the Spanish MCYT and the EU-SOL European Union Integrated Project. Support by the JSPS and CSIC under the Japan–Spain research cooperative programme and Grants-in-Aid for Scientific Research on Priority Areas of the MEXT of Japan are also acknowledged. C.N. was recipient of an I3P postdoctoral contract from the CSIC.

Author Contributions C.N., J.A.A., E.C.-O., C.A.C., J.S. and S.P. performed experiments and analysed the results. S.T. and K.S. provided the rice Hd3a construct and performed the Hd3a protein mobility studies. S.P. designed the experiments and wrote the manuscript.

Author Information Microarray data on StSP6Aox, StSP6A RNAi and stolon gene expression during photoperiod induction have been deposited in the Gene Expression Omnibus under the accession codes GSE31178 and GSE31336. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.P. (sprat@cnb.csic.es).

METHODS

Plant materials and growth conditions. Andigena 7540 wild-type lines and the antisense phytochrome B²¹ (*phyB*) and AtCO overexpressor¹⁷ (AtCOox) lines were grown in the greenhouse under long-day conditions. Growth chambers were used for controlled short-day (8 h light, 16 h dark), short-day/night-break (short day plus a 30-min pulse of light given 8 h after the beginning of the dark period) and long-day (16 h light/8 h dark) treatments.

Transgenic Andigena plants carrying the *rolC::Hd3a-GFP*¹⁸, *StSP6A* overexpression, *StSP6A* RNAi, *StSP3D* RNAi, *StCOox* and *StCO* RNAi constructs were generated by *Agrobacterium*-mediated transformation of leaf explants as described previously¹⁷. Different graft combinations were obtained as previously described¹⁷ and their tuberization phenotype was scored under long-day conditions. Tuber formation in *StSP6A* RNAi and *StSP3D* RNAi lines was analysed by growing the plants under long-day greenhouse conditions until a ten-leaf stage, before transferring them to short-day conditions. For flowering studies, plants were grown under high-irradiance ($>200 \mu\text{E s}^{-1} \text{m}^{-2}$) long-day conditions. At least nine replicates of each line were used for these studies.

Arabidopsis constans-1 (co-1) and *flowering locus t-1 (ft-1)* mutants in the Columbia (Col-0) and Landsberg erecta (Ler) backgrounds were transformed using the floral dip method. Flowering time was measured as the number of rosette leaves at floral initiation under long-day conditions, in at least ten individuals.

Plasmid constructs. The *rolC::Hd3a-GFP* construct has been described elsewhere¹⁸. Transcripts corresponding to the *StSP6A*, *StSP3D* and *StCO* genes were amplified using primers designed on the tomato sequences (AY186737, AY186735 and AY490253, respectively). The *StSP6A* overexpression construct was generated by amplifying the protein-coding region and then inserting the PCR product into the *Sma*I site of the pBinAR binary vector, between the 35S promoter and the *ocs* terminator. To silence the *StSP6A*, *StSP3D* and *StCO* transcripts, Andigena plants were transformed with RNAi constructs designed on the non-conserved region of the genes. The amplification products were cloned into the pENTR/D-TOPO plasmid (Invitrogen) and inserted in opposite orientations by recombination with the LR Clonase II enzyme (Gateway Technology, Invitrogen) into the pBIN19RNAi destination vector. The pBIN19RNAi interference vector was generated by partial *Xba*I/*Hind*III digestion of the pH7GWGW2(II) plasmid (<http://gateway.psb.ugent.be/>) and insertion of the Gateway cassette fragment into the same restriction sites of the pBIN19 binary vector. The *StCO* overexpression construct was generated by amplifying and cloning the protein-coding region into the pENTR/D-TOPO plasmid (Invitrogen), and further insertion by recombination into the pGW2 destination vector³¹. To generate the *StSP6A* ethanol-inducible lines, the pBinSRNA-GW destination vector was created by insertion of the blunt-ended pAlcA-R1-R2-t35S cassette from the AlcAP-GW pGreen vector (gift of Patrick Laufs, INRA), generated by *Xba*I digestion, into the blunted *Hind*III sites of the binary vector pBinSRNACatN. The *StSP6A* coding region was then introduced in this destination vector by LR Clonase recombination. A list of the primer sets used to generate these constructs is shown in Supplementary Table 3.

Real-time and semiquantitative RT-PCR analyses. Expression of potato *FT*- and *TFL1*-like family members was quantified by real-time PCR. First-strand complementary DNA was synthesized from 2 μg total RNA and 1 μl of the reaction used for real-time gene expression analysis with the SYBR Green PCR master mix (Applied Biosystems). Quantitative PCR was performed using the Power SYBR Green PCR Master Mix (Applied Biosystems) on an ABI7500 Real-Time PCR System (Applied Biosystems), following the manufacturer's instructions. Primer pairs were specifically designed for each gene using PRIMER EXPRESS 3.0 software (Applied Biosystems) and probed for high-efficiency amplification under standard quantitative PCR conditions. All reactions were carried out at least in two independent biological replicates. The Pfaffl³² method was used for relative quantification of gene expression. Direct $2^{-\Delta C_T}$, where $\Delta C_T = C_T$ (target gene) $- C_T$ (actin) (ref. 33), was used to generate the tissue-specific heat map in Fig. 1e. The comparative critical threshold ($2^{-\Delta\Delta C_T}$) method³⁴ was used to analyse relative *Hd3a* and *StSP6A* expression levels (Supplementary Figs 1 and 10).

Identical procedures were used for semiquantitative RT-PCR, except that amplification was conducted in a Peltier thermal cycler (PTC-200, MJ Research) and one-tenth of the complementary DNA reaction was used for actin amplification. For sets of specific primers and product lengths, see Supplementary Table 3.

Microarray sample hybridization and analysis. Stolons of wild-type plants were collected in long days and two, six and eight days after transferring the plants to

short days. Time-course profiling analyses of short-day-induced stolons were performed as described previously³⁵.

Arrays of *StSP6A* overexpression versus wild type (sampled in long days) and *StSP6A* RNAi stolons versus wild type (sampled in short days, day 6) were performed with samples from three independent lines. Samples were hybridized against POCI (Potato Oligo Chip Initiative) microarrays, and background correction and normalization were performed using LIMMA^{36,37}. The obtained data was statistically checked (false discovery rate, <0.05) and genes with a log ratio change of +2 or -2 for *StSP6A* RNAi and +1.8 or -1.8 for *StSP6Aox* plants were selected. The hierarchical cluster of genes present in both experiments was calculated, compared with the short-day profile and represented using the TIGR MEV free software (<http://www.tm4.org/mev/>). Genes found to be differentially expressed are listed in Supplementary Table 1.

Immunoblot analysis. Total protein extracts from stolon tissues were obtained in lysis buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1% Na deoxycholate, 0.5% Triton X-100, 1 mM PMSF and protease inhibitors). For analysis of protein graft transmissibility, total extracts were incubated overnight with 10 μl of an anti-GFP affinity matrix (MBL). After extensive washing, unbound and bound proteins were separated by 10% SDS-polyacrylamide gel electrophoresis, blotted onto nitrocellulose and probed with an anti-GFP antibody (Roche). Immunoreactive proteins were detected with the SuperSignal West Pico Chemiluminescence kit (Pierce).

Microscopy. GFP fluorescence was observed on longitudinal stolon sections (150 μm) obtained with a vibratome (PELCO 101). Fluorescence was excited with a 488-nm argon laser and emission images were collected in the 500–600-nm range using a Leica TCS SP5 spectral confocal microscope.

For scanning electron microscopy, potato stolons and apical meristems were frozen in an Oxford CT 1500 cryosystem (Oxford Instruments), sublimated under vacuum and observed using a DIOL JSM 5410 electronic microscope operating at 10 kV.

Phylogenetic analysis. Full-length protein sequences, except sequences of the potato *FT*- and *TFL1*-like family members, were obtained from GenBank. Sequences of potato *FT*- and *TFL1*-like family members were obtained from the Potato Genome Sequencing Consortium Data Release²⁰ (<http://potatogenomics.plantbiology.msu.edu/>) by TBLASTX using as a query previously described tomato sequences¹⁹ (*SlSP*, no. U84140; *SlSP6A*, no. AY186737; *SlSP3D*, no. AY186735; *SlSP9D*, no. AY186738; *SlSP2I*, no. AY186734; and *SlSP5G*, no. AY186736). The best genome matches (Supplementary Table 2) were downloaded and open reading frames were predicted using FGENESH (<http://linux1.softberry.com/berry.phtml>) and ORF Finder (<http://www.ncbi.nlm.nih.gov/projects/gorf/>). A counterpart to *SlSP2I* was not found in the potato genome. Phylogenetic analyses were conducted using MEGA4³⁸. Sequences were aligned with the COBALT program (<http://www.ncbi.nlm.nih.gov/tools/cobalt/>) and their evolutionary relationship inferred using the neighbour-joining method³⁹. All ambiguous positions were removed for each sequence pair. The evolutionary distances were computed using the Poisson correction method and the bootstrap consensus tree was inferred from 1,000 replicates. The accession numbers for the corresponding genes are indicated in the tree (Supplementary Fig. 2).

- Nakagawa, T. *et al.* Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *J. Biosci. Bioeng.* **104**, 34–41 (2007).
- Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**, e45 (2001).
- Schmittgen, T. D. & Livak, K. J. Analyzing real-time PCR data by the comparative C_T method. *Nature Protocols* **3**, 1101–1108 (2008).
- Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method. *Methods* **25**, 402–408 (2001).
- Kloosterman, B. *et al.* Genes driving potato tuber initiation and growth: identification based on transcriptional changes using the POCI array. *Funct. Integr. Genomics* **8**, 329–340 (2008).
- Smyth, G. K. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3**, 3 (2004).
- Smyth, G. K. & Speed, T. Normalization of cDNA microarray data. *Methods* **31**, 265–273 (2003).
- Tamura, K., Dudley, J., Nei, M. & Kumar, S. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**, 1596–1599 (2007).
- Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).

Structure of human mitochondrial RNA polymerase

Rieke Ringel¹, Marina Sologub², Yaroslav I. Morozov², Dmitry Litonin^{1,2}, Patrick Cramer¹ & Dmitry Temiakov²

Transcription of the mitochondrial genome is performed by a single-subunit RNA polymerase (mtRNAP) that is distantly related to the RNAP of bacteriophage T7, the pol I family of DNA polymerases, and single-subunit RNAPs from chloroplasts^{1–4}. Whereas T7 RNAP can initiate transcription by itself, mtRNAP requires the factors TFAM and TFB2M for binding and melting promoter DNA^{5–7}. TFAM is an abundant protein that binds and bends promoter DNA 15–40 base pairs upstream of the transcription start site, and stimulates the recruitment of mtRNAP and TFB2M to the promoter^{8,9}. TFB2M assists mtRNAP in promoter melting and reaches the active site of mtRNAP to interact with the first base pair of the RNA–DNA hybrid¹⁰. Here we report the X-ray structure of human mtRNAP at 2.5 Å resolution, which reveals a T7-like catalytic carboxy-terminal domain, an amino-terminal domain that remotely resembles the T7 promoter-binding domain, a novel pentatricopeptide repeat domain, and a flexible N-terminal extension. The pentatricopeptide repeat domain sequesters an AT-rich recognition loop, which binds promoter DNA in T7 RNAP, probably explaining the need for TFAM during promoter binding. Consistent with this, substitution of a conserved arginine residue

in the AT-rich recognition loop, or release of this loop by deletion of the N-terminal part of mtRNAP, had no effect on transcription. The fingers domain and the intercalating hairpin, which melts DNA in phage RNAPs, are repositioned, explaining the need for TFB2M during promoter melting. Our results provide a new venue for the mechanistic analysis of mitochondrial transcription. They also indicate how an early phage-like mtRNAP lost functions in promoter binding and melting, which were provided by initiation factors *in trans* during evolution, to enable mitochondrial gene regulation and the adaptation of mitochondrial function to changes in the environment.

We crystallized a fully functional variant of a recombinant human mtRNAP (residues 105–1230) that requires the presence of both TFAM and TFB2M for efficient transcription initiation on a double-stranded promoter DNA (Supplementary Fig. 1). The structure was determined at 2.5 Å resolution by molecular replacement with the use of a truncated T7 RNAP structure as a search model², and was refined to a free *R*-factor of 0.23 (Methods, and Supplementary Table 1).

The mtRNAP structure has the shape of a right hand with palm, fingers and thumb subdomains, characteristic of the pol A family of

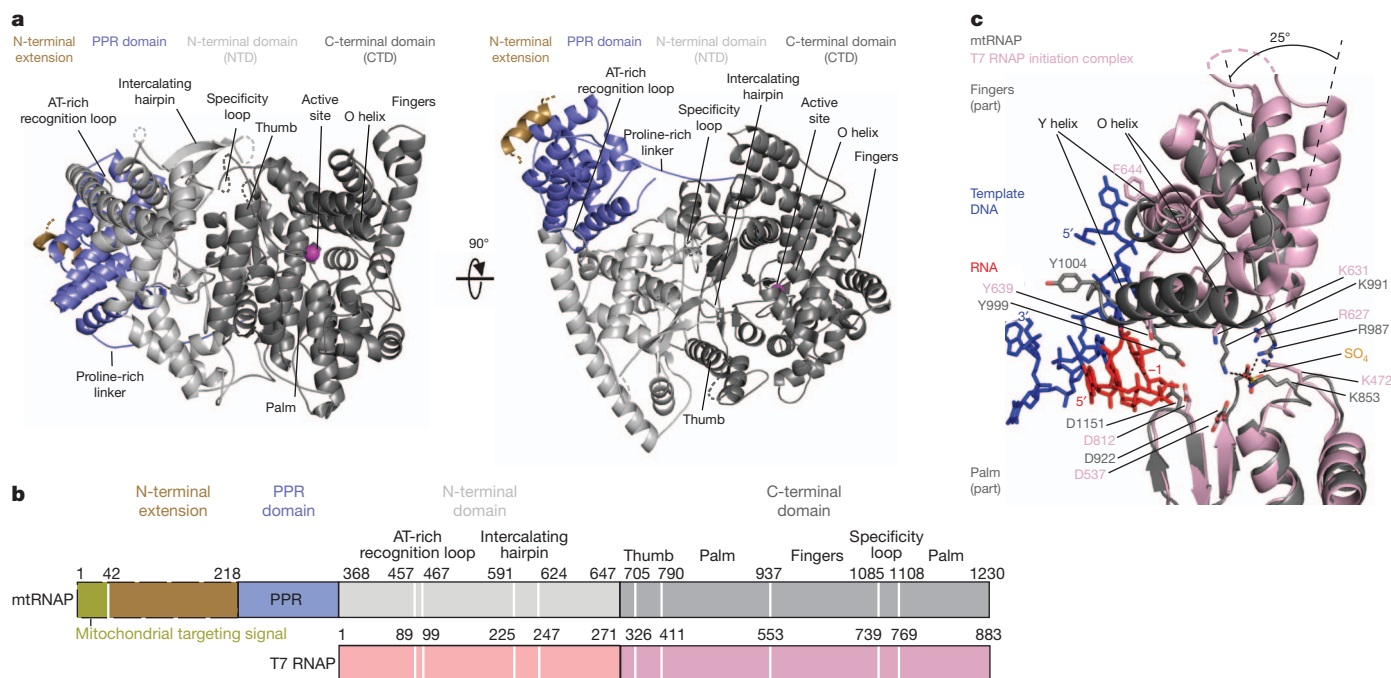


Figure 1 | Crystal structure of human mtRNAP. **a**, Two views of a ribbon model with the major domains and structural elements indicated. The CTD that is conserved in all single-stranded RNAPs is in dark grey, the NTD in silver, the PPR domain in blue, and the N-terminal extension helix in sand. The active site is indicated by a magenta sphere for a modelled catalytic metal ion. **b**, Schematic comparison of mtRNAP with T7 (PDB 1QLN) RNAP. Prominent structural elements are indicated. mtRNAP-specific residues 1–368 include the mitochondrial targeting signal, the N-terminal extension and the PPR domain.

Regions in mtRNAP that are not visible in the crystal structure include residues 592–602, 736–769 and 1086–1105. **c**, Superimposition of the active-centre regions of mtRNAP (grey) and the T7 RNAP initiation complex (PDB 1QLN, light pink) reveals a clenched conformation due to a roughly 25° rotation of the fingers subdomain around the O-helix axis. Template DNA is in blue, RNA in red. Conserved catalytic residues, residues implicated in NTP binding, and a sulphate ion are shown with sticks.

¹Gene Center and Department of Biochemistry, Center for Integrated Protein Science Munich (CIPSM), Ludwig-Maximilians-Universität München, Feodor-Lynen-Strasse 25, 81377 Munich, Germany.

²Department of Cell Biology, University of Medicine and Dentistry of New Jersey, School of Osteopathic Medicine, 2 Medical Center Dr, Stratford, New Jersey 08084, USA.

nucleotidyltransferases (Fig. 1). We use here the canonical polymerase domain nomenclature and the previously defined names for functional elements in T7 RNAP^{2,11}, even though these have functional implications that are not borne out in mtRNAP. The highly conserved palm in the C-terminal domain (CTD) superimposes well on the corresponding palm in T7 RNAP (root mean squared deviation 1.0 Å, 121 backbone atoms). The marked conservation of the active centre, including the O helix, suggests that the mechanisms of substrate binding and selection are conserved between mtRNAP and T7 RNAP^{12,13}. The O helix binds a sulphate ion with its residues R987, K991 and K853 (Fig. 1c) near the position of the phosphate groups of an incoming nucleoside triphosphate (NTP) in the T7 RNAP elongation complex^{12,13}. A portion of the thumb (residues 736–769) and the flexible specificity loop (residues 1086–1105) are not visible in the structure, but adjacent elements occupy almost identical positions in T7 RNAP, indicating that these elements have similar orientations and functions.

The most notable difference between the CTD of T7 RNAP and mtRNAP is the position of the fingers subdomain, which is rotated as a rigid body by roughly 25° approximately around the axis of the O helix, accompanied by a tilting of the O and Y helices (Fig. 1c). These movements result in a 15 Å translation of the N-terminal part of the Y helix towards the promoter-binding domain (PBD). The observed position of the fingers domain is distinct from the previously observed positions in the ‘closed’ elongation-complex structure of T7 RNAP, in which the O helix is positioned to deliver the NTP into the active site¹² (Fig. 1c). We refer to this novel RNAP conformation as ‘clenched’. The clenched conformation is unlikely to occur during the nucleotide addition cycle, because the entrance to the active site may be occluded by the conserved Y-helix residue Y1004, which was implicated in downstream duplex melting^{14,15} and is partly mobile in our structure. It is unclear whether the clenched conformation is functional or is a result of crystal packing.

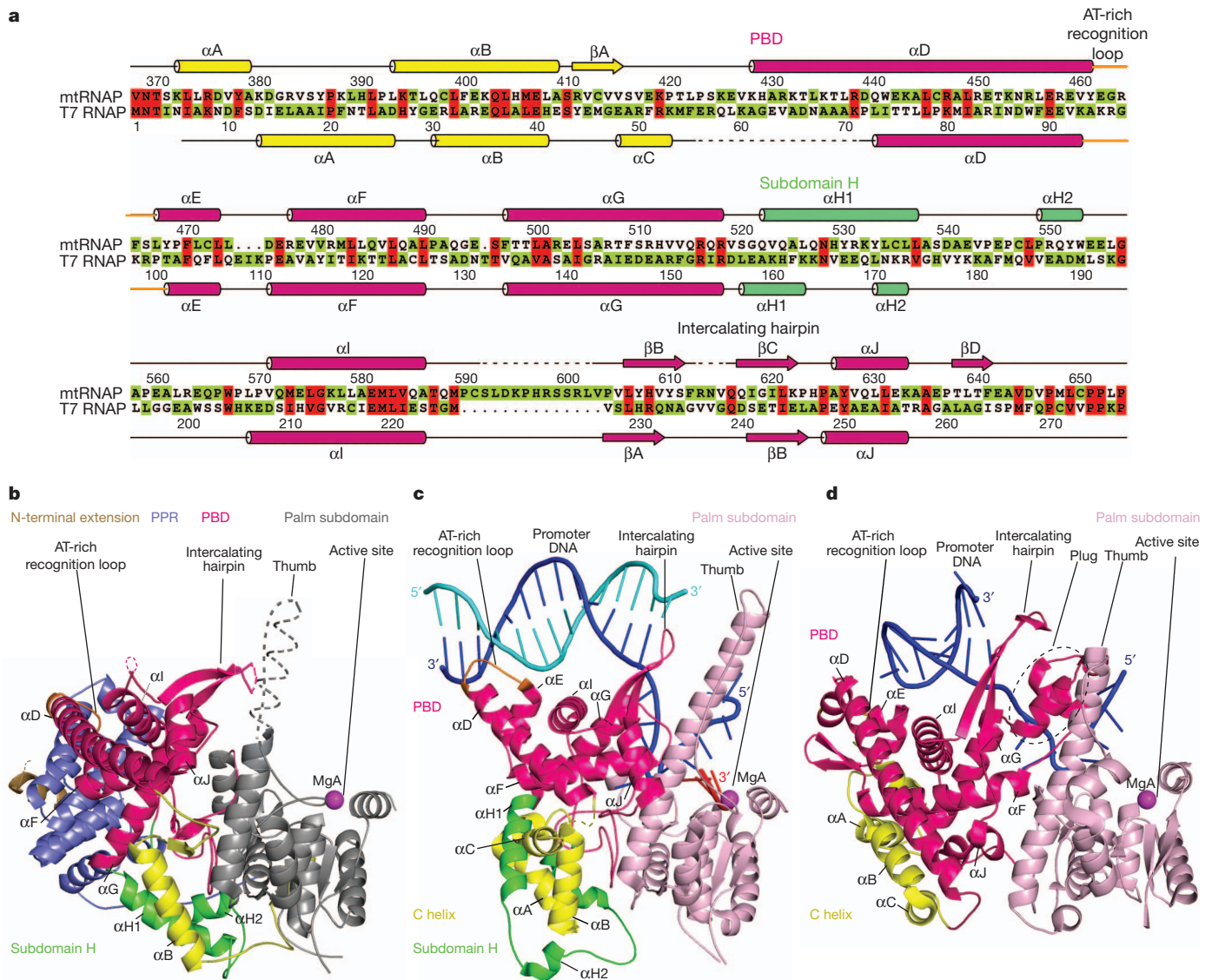


Figure 2 | Variation in RNAP N-terminal domains. **a**, Sequence alignment and structural conservation of human mtRNAP and T7 RNAP (PDB 1QLN). Secondary-structure elements are indicated above and below the sequences for mtRNAP (residues 368–654) and T7 RNAP, respectively (cylinders, α -helices; arrows, β -strands; lines, loops). The PBD is in pink, the C helix in yellow, and subdomain H in green. The AT-rich recognition loop (orange) and the intercalating hairpin are indicated. Identical residues are highlighted in red. Dashed lines indicate regions that are not visible in the mtRNAP structure.

b–d, Comparison of the NTDs of mtRNAP (b) with initiation complexes of T7 RNAP (PDB 1QLN) (c) and N4 RNAP (PDB 3Q24) (d). Structures are aligned with respect to their conserved palm subdomains. Colour code and residue borders are as in a. RNA is in red, template DNA in blue, and a non-template strand in cyan. Of the CTD (grey or light pink) only the palm subdomain (residues 790–831, 911–937 and 1125–1178) and the thumb (residues 677–790) are shown.

The large N-terminal region of mtRNAP (residues 1–647; Fig. 1b) shares no sequence homology with T7 RNAP. Nevertheless, the C-terminal part of this region (residues 369–647) is structurally similar to the N-terminal domain of phage RNAPs, and we therefore refer to it as the N-terminal domain (NTD). In particular, the six-helix bundle of the PBD in phage RNAPs, which includes helices D, E, F, G, I and J, has a counterpart in mtRNAP (Fig. 2). However, there are tangible changes in the orientation of the PBD helices between mtRNAP and T7 RNAP (root mean squared deviation 9 Å; 109 backbone atoms). In comparison with T7 RNAP, the N-terminal part of the PBD (residues 425–519) is rotated by 32°, whereas the C-terminal part (residues 567–654) is rotated by 24° in mtRNAP.

The structure further reveals the loops corresponding to the intercalating hairpin and the AT-rich recognition loop of T7 RNAP. The hairpin connects the conserved helices I and J and has a 14-residue insertion that is partly visible in the mtRNAP structure (Fig. 2a). The loop that connects helices D and E of the PBD corresponds to the AT-rich recognition loop but is shorter, whereas the D helix is extended by comparison with T7 RNAP (Fig. 2b, c). Another significant difference in the NTD of mtRNAP is the alternative fold of the region at the position of subdomain H in T7 RNAP. It is likely that this region of the mtRNAP NTD undergoes refolding during the transition from initiation to elongation, to form a part of an RNA exit pore, as described for T7 RNAP^{11,14,15}. It remains to be seen whether the observed conformation is specific for mtRNAP or whether it also occurs in T7 RNAP but has not yet been observed. In phage N4 RNAP, which is distantly related to T7 RNAP, subdomain H is absent (Fig. 2d).

The NTD of mtRNAP is connected to a unique helical domain (residues 218–368) through an extended, nearly linear, proline-rich linker (Figs 1 and 3). This domain consists of nine α -helices, of which four comprise two pentatricopeptide repeat (PPR) motifs found in plant and mitochondrial proteins^{16–18}. We refer to this domain as the PPR domain. PPR-motif-containing proteins are apparently involved in RNA editing and processing^{19,20}, but their structures are unknown. Our structure of the PPR domain shows that the conserved residues of the PPR motif form the hydrophobic core of a helix–turn–helix fold and the interface between antiparallel α -helices (α D' and α E', α E' and α F', α F' and α G') and parallel helices (α E' and α G')

(Fig. 3a, b). The PPR domain interacts tightly with the NTD. Helices α A' and α B' form a large hydrophobic interface (about 950 Å²) that includes the tip of the AT-rich recognition loop and its flanking helices α E and α D (Fig. 3c). In addition, a salt bridge between R225 and E574 and numerous hydrogen bonds stabilize the PPR–NTD interface (Fig. 3c).

The region located N-terminal of the PPR domain, called here the N-terminal extension, is mostly mobile, except for a single unassigned α -helix that docks to a hydrophobic pocket formed by α F' and α G' of the PPR domain (Fig. 3b). The N-terminal extension of human mtRNAP is required for function, because deletion of the N-terminal 200 residues results in a truncated polymerase that is catalytically active but unable to initiate promoter-directed transcription (Supplementary Fig. 1).

Footprinting experiments suggested that during transcription initiation mtRNAP (in complex with TFB2M) and T7 RNAP occupy 14 and 17 base pairs, respectively, of upstream promoter DNA^{7,10} (Fig. 4a). The rest of the upstream promoter region in mammalian mtDNA (–35 to –15) is covered by TFAM⁷, suggesting that mtRNAP and T7 RNAP use different modes of promoter binding. To explore this we modelled the putative trajectory of upstream promoter DNA in mtRNAP by superimposing the palm subdomains of mtRNAP and the T7 RNAP initiation complex²¹ (Fig. 4b). In the resulting model, which remains tentative because DNA binding can cause structural changes, upstream promoter DNA runs along the NTD, and the template single strand descends into the mtRNAP active site (Figs 1c and 4b). The tip of the specificity loop may reach into the major groove of promoter DNA and read the sequence between –6 and –11 base pairs while its base may interact with the phosphodiester backbone of a single-stranded DNA, essentially as in phage RNAPs^{21,22}. This is consistent with a functional analysis of point mutations in the specificity loop of yeast mtRNAP²³.

Modelling also reveals structural differences in mtRNAP, relative to T7 RNAP, that may render mtRNAP reliant on the initiation factor TFB2M and responsive to activation by TFAM. In particular, the AT-rich recognition loop has distinct structures and functions in the two systems. In comparison with T7 RNAP, helix D is rotated away from the minor groove of DNA by 45° and the tip of the AT-rich recognition loop is shifted by 22 Å (Fig. 4c) and sequestered by

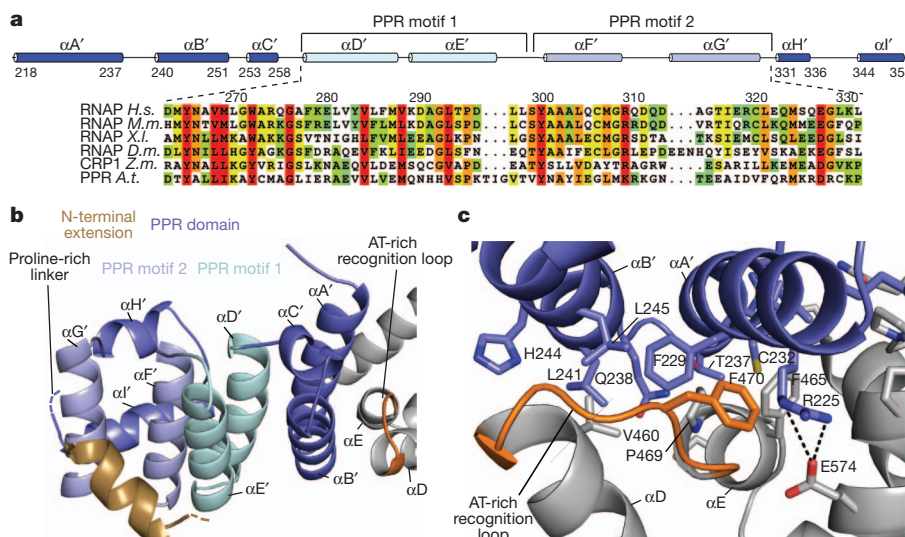


Figure 3 | PPR domain. **a**, Sequence alignment of human mtRNAP PPR domain, including PPR motifs 1 and 2, with predicted PPR motifs in mtRNAPs of different species (lanes 2–4), and chloroplast-targeting proteins of plants (lines 5,6). Abbreviations: *H.s.*, *Homo sapiens*; *M.m.*, *Mus musculus*; *X.l.*, *Xenopus laevis*; *D.m.*, *Drosophila melanogaster*; *Z.m.*, *Zea mays*; *A.t.*, *Arabidopsis thaliana*. Residues that are identical, highly conserved, conserved, and distinct, are highlighted in red, orange, yellow, and green, respectively.

b, Structure of PPR domain (blue) with PPR motifs 1 (light cyan) and 2 (light blue). Each PPR motif shows a helix–turn–helix fold. Helices α A' and α B' are adjacent to the NTD (silver) and interact with the AT-rich recognition loop (orange). PPR motif 1 interacts with a helix in the N-terminal extension (sand). **c**, Interface between PPR domain and NTD. Colours are as in **a**. Interface residues are shown in stick representation.

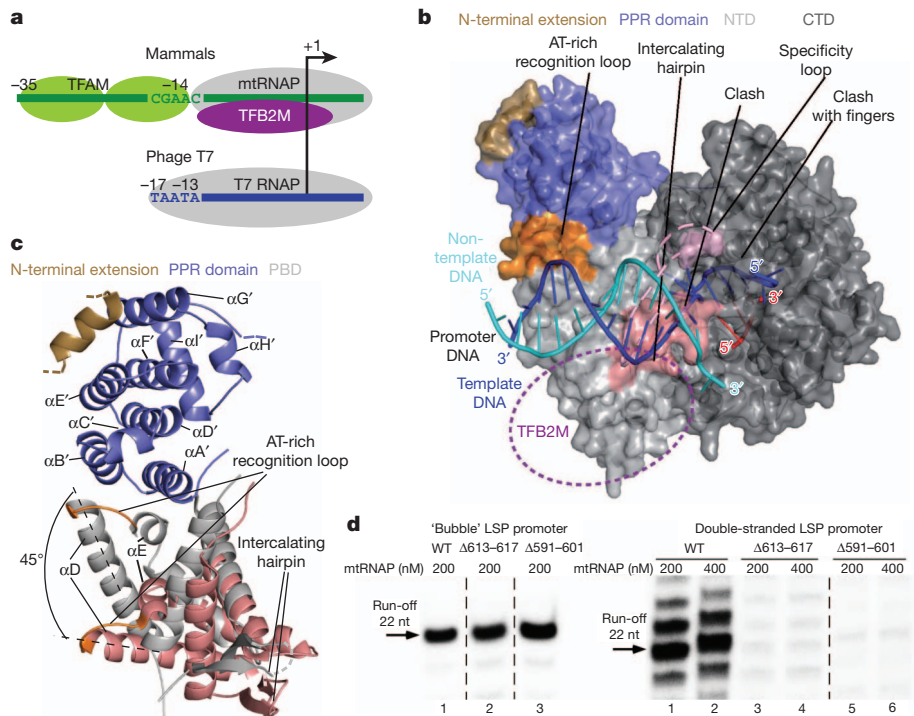


Figure 4 | Promoter binding and melting by mtRNAP. **a**, Differences in topology of the initiation complexes of T7 and mammalian mitochondrial transcription machineries, based on foot-printing data and functional assays. **b**, Modelled location of upstream promoter DNA on mtRNAP, based on a superimposition with the T7 RNAP initiation complex (PDB 1QLN). mtRNAP is depicted as a molecular surface. Loops in mtRNAP that interact with DNA in the T7 system are indicated. The portion of the specificity loop not visible in the structure is modelled based on the corresponding T7 loop (dashed line). The AT-rich recognition loop is buried by the PPR domain (blue) and not seen. The intercalating hairpin (rose) is shifted by 7 Å compared with the T7 RNAP initiation complex, and clashes with the template DNA strand (blue). A clash is also observed between the fingers subdomain and the +1 template base. The presumed binding site of TFB2M is indicated by a dash oval. **c**, Relative

extensive interactions with the PPR domain (Fig. 3c). Consistent with this, a mtRNAP variant that lacks the PPR domain, and should contain a free AT-rich recognition loop, cannot initiate transcription (Supplementary Fig. 1). In addition, the positively charged residues important for promoter interactions by T7 RNAP are not conserved in the mtRNAP loop, and mutations R458A/R464A have no effect on transcription activity (Supplementary Fig. 2). These results indicate that human mtRNAP does not use its AT-rich recognition loop for promoter binding, explaining why it depends on transcription factors for initiation. The AT-rich recognition loop also has a different structure and function in RNAP of phage N4, where it recognizes a hairpin promoter²² (Fig. 2d).

mtRNAP shows two additional structural differences relative to T7 RNAP that explain why it can not melt DNA and requires TFB2M for promoter melting. First, the position of the intercalating hairpin is not compatible with promoter melting as observed in T7 RNAP (Fig. 2b, c). The tip of the loop is translated 7 Å away from its position in T7 and clashes with the template strand of DNA (Fig. 4b). Second, a 25° rotation of the fingers domain towards the NTD (Fig. 1c) blocks access of the single-stranded DNA to the active site.

Two mechanisms of promoter melting may be considered. First, binding of TFB2M may reposition the intercalating hairpin so that it can function as in T7 and N4 RNAP^{21,22}. Second, DNA melting may be achieved not by the intercalating hairpin but by an unidentified structural element in TFB2M. To investigate this we constructed variants of mtRNAP lacking either five residues at the tip of the intercalating hairpin (Δ613–617) or lacking a preceding insertion (Δ591–601).

orientation of the PBDs of mtRNAP (silver) and T7 RNAP initiation complex (PDB 1QLN, salmon) based on homology modelling of the corresponding palm subdomain. The AT-rich recognition loop (orange) is shifted and rotated by 45°. **d**, An intercalating hairpin region is required for promoter melting in mtRNAP. Transcription run-off assays with wild-type (WT) and mutant RNAPs were performed with bubble (left panel) and double-stranded (right panel) light-strand promoter templates for 30 min at 35 °C. Transcription factors TFAM (50 nM) and TFB2M (150 nM) were added to reactions that involved double-stranded template (right panel), whereas transcription with the bubble template was factor-independent (left panel). Reaction products were resolved with 20% PAGE containing 6 M urea, and analysed with a PhosphorImager. nt, nucleotides.

These mutant polymerases had an activity similar to that of the wild-type enzyme on a pre-melted 'bubble' light-strand promoter template (Fig. 4d) but were essentially inactive on a double-stranded light-strand promoter template, suggesting a role of the intercalating hairpin in DNA melting (Fig. 4d). To explore this further, we built a model of a mitochondrial transcription initiation complex, which takes available structural²⁴, biochemical and genetic data^{10,25–27} into account (Supplementary Fig. 3). The model remains tentative but indicates that TFB2M may insert between the intercalating hairpin and thumb subdomain, to reposition the intercalating hairpin and the fingers, such that they function in promoter melting (Supplementary Fig. 4). TFB2M may also reach near the active site to stabilize an early transcribing complex by binding the template strand and incoming NTP as suggested¹⁰.

METHODS SUMMARY

Human mtRNAP, TFB2M and TFAM were expressed and purified as described¹⁰. The coding sequences of human mtRNAP N-terminal deletion mutants Δ104 (residues 105–1230), Δ200 (201–1230) and Δ368 (369–1230) were amplified by PCR and cloned into vector pProEx(Hb) (Invitrogen), to express N-terminally hexahistidine-tagged proteins. Deletion mutants Δ613–617 and Δ591–601GG and substitution mutant R458A/R464A were obtained by site-directed mutagenesis (Quick-change; Stratagene). Transcription run-off reactions were performed with synthetic DNA templates as described¹⁰ (Supplementary Fig. 1). For crystallization, Δ104 mtRNAP was treated *in situ* with ArgC protease (1:1,000 w/w) for 1–2 h at 23 °C. Initial crystals were obtained in hanging drops over 12% poly(ethylene glycol) (PEG) 4000, 10% glycerol, 30 mM MES, 60 mM (NH₄)₂SO₄, 40 mM Tris-HCl pH 8.0, 20 mM dithiothreitol at 20 °C, and used for micro-seeding, which

produced rod-shaped crystals of up to $0.3 \times 0.1 \times 0.1 \text{ mm}^3$ in size. Diffraction data were collected in 0.5° increments at 100 K on a MarCCD detector at the Swiss Light Source in Villigen, Switzerland (Supplementary Table 1). Data were integrated with MOSFLM²⁸ and scaled with SCALA²⁹. The structure was solved by molecular replacement using the T7 RNAP structure² (PDB 1QLN) truncated to polyaniline and reduced to residues 246–354, 392–568, 691–741 and 766–883. The replacement solution was subjected to rigid-body refinement with Refmac-5 (ref. 30). The model was iteratively built with COOT and refined with autoBuster (Global Phasing) to a final free *R*-factor of 23%. The refined structure includes residues 218–591, 603–611, 616–735, 770–1085 and 1109–1230. Residues 326–360 and the helix in the N-terminal extension were built as polyaniline because of poor side-chain density. Figures were prepared with Pymol (Schrödinger).

Received 9 May; accepted 10 August 2011.

Published online 25 September 2011.

- Masters, B. S., Stohl, L. L. & Clayton, D. A. Yeast mitochondrial RNA polymerase is homologous to those encoded by bacteriophages T3 and T7. *Cell* **51**, 89–99 (1987).
- Cheetham, G. M. & Steitz, T. A. Structure of a transcribing T7 RNA polymerase initiation complex. *Science* **286**, 2305–2309 (1999).
- Gaspari, M., Larsson, N. G. & Gustafsson, C. M. The transcription machinery in mammalian mitochondria. *Biochim. Biophys. Acta* **1659**, 148–152 (2004).
- Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A. & Steitz, T. A. Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* **256**, 1783–1790 (1992).
- Litonin, D. et al. Human mitochondrial transcription revisited: only TFAM and TFB2M are required for transcription of the mitochondrial genes *in vitro*. *J. Biol. Chem.* **285**, 18129–18133 (2010).
- Falkenberg, M. et al. Mitochondrial transcription factors B1 and B2 activate transcription of human mtDNA. *Nature Genet.* **31**, 289–294 (2002).
- Gaspari, M., Falkenberg, M., Larsson, N. G. & Gustafsson, C. M. The mitochondrial RNA polymerase contributes critically to promoter specificity in mammalian cells. *EMBO J.* **23**, 4606–4614 (2004).
- Ekstrand, M. I. et al. Mitochondrial transcription factor A regulates mtDNA copy number in mammals. *Hum. Mol. Genet.* **13**, 935–944 (2004).
- Dairaghi, D. J., Shadel, G. S. & Clayton, D. A. Human mitochondrial transcription factor A and promoter spacing integrity are required for transcription initiation. *Biochim. Biophys. Acta* **1271**, 127–134 (1995).
- Sologub, M., Litonin, D., Anikin, M., Mustaev, A. & Temiakov, D. TFB2 is a transient component of the catalytic site of the human mitochondrial RNA polymerase. *Cell* **139**, 934–944 (2009).
- Durniak, K. J., Bailey, S. & Steitz, T. A. The structure of a transcribing T7 RNA polymerase in transition from initiation to elongation. *Science* **322**, 553–557 (2008).
- Yin, Y. W. & Steitz, T. A. The structural mechanism of translocation and helicase activity in T7 RNA polymerase. *Cell* **116**, 393–404 (2004).
- Temiakov, D. et al. Structural basis for substrate selection by T7 RNA polymerase. *Cell* **116**, 381–391 (2004).
- Tahirov, T. H. et al. Structure of a T7 RNA polymerase elongation complex at 2.9 Å resolution. *Nature* **420**, 43–50 (2002).
- Yin, Y. W. & Steitz, T. A. Structural basis for the transition from initiation to elongation transcription in T7 RNA polymerase. *Science* **298**, 1387–1395 (2002).
- Rodeheffer, M. S., Boone, B. E., Bryan, A. C. & Shadel, G. S. Nam1p, a protein involved in RNA processing and translation, is coupled to transcription through an interaction with yeast mitochondrial RNA polymerase. *J. Biol. Chem.* **276**, 8616–8622 (2001).
- Small, I. D. & Peeters, N. The PPR motif—a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem. Sci.* **25**, 46–47 (2000).
- Lightowlers, R. N. & Chrzanowska-Lightowlers, Z. M. PPR (pentatricopeptide repeat) proteins in mammals: important aids to mitochondrial gene expression. *Biochem. J.* **416**, e5–e6 (2008).
- Hammani, K. et al. The pentatricopeptide repeat protein OTP87 is essential for RNA editing of *nad7* and *atp1* transcripts in *Arabidopsis* mitochondria. *J. Biol. Chem.* **286**, 21361–21371 (2011).
- Aphasizheva, I., Maslov, D., Wang, X., Huang, L. & Aphasizhev, R. Pentatricopeptide repeat proteins stimulate mRNA adenylation/uridylation to activate mitochondrial translation in trypanosomes. *Mol. Cell* **42**, 106–117 (2011).
- Cheetham, G. M., Jeruzalmi, D. & Steitz, T. A. Structural basis for initiation of transcription from an RNA polymerase–promoter complex. *Nature* **399**, 80–83 (1999).
- Gleghorn, M. L., Davydova, E. K., Rothman-Denes, L. B. & Murakami, K. S. Structural basis for DNA-hairpin promoter recognition by the bacteriophage N4 virion RNA polymerase. *Mol. Cell* **32**, 707–717 (2008).
- Nayak, D., Guo, Q. & Sousa, R. A promoter recognition mechanism common to yeast mitochondrial and phage T7 RNA polymerases. *J. Biol. Chem.* **284**, 13641–13647 (2009).
- Schubot, F. D. et al. Crystal structure of the transcription factor sc-mtTFB offers insights into mitochondrial transcription. *Protein Sci.* **10**, 1980–1988 (2001).
- Mangus, D. A., Jang, S. H. & Jaehning, J. A. Release of the yeast mitochondrial RNA polymerase specificity factor from transcription complexes. *J. Biol. Chem.* **269**, 26568–26574 (1994).
- Clifton, P. F., Park, J. Y., Davis, B. P., Jang, S. H. & Jaehning, J. A. Identification of three regions essential for interaction between a sigma-like factor and core RNA polymerase. *Genes Dev.* **11**, 2897–2909 (1997).
- Savkina, M., Temiakov, D., McAllister, W. T. & Anikin, M. Multiple functions of yeast mitochondrial transcription factor Mtf1p during initiation. *J. Biol. Chem.* **285**, 3957–3964 (2010).
- Leslie, A. G. W. *Joint CCP4 and ESF-EACMB Newsletter Protein Crystallogr.* No. 26 (Daresbury Laboratory, 1992).
- Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72–82 (2006).
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Cheung and D. Kostrewa for help with crystallography, and W. T. McAllister and M. Anikin for the critical reading of the manuscript. We acknowledge the crystallization facility at the Max Planck Institute of Biochemistry, Martinsried. Part of this work was performed at the Swiss Light Source at the Paul Scherrer Institut, Villigen, Switzerland. Use of the Advanced Photon Source at Argonne National Laboratory was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, under contract no. DE-AC02-06CH11357. Use of the Lilly Research Laboratories Collaborative Access Team (LRL-CAT) beamline at Sector 31 of the Advanced Photon Source was provided by Eli Lilly Company, which operates the facility. P.C. was supported by the Deutsche Forschungsgemeinschaft, SFB646, TR5, FOR1068, the Nanosystems Initiative Munich NIM, an Advanced Investigator Grant from the European Research Council ERC, and the Jung-Stiftung. D.T. was supported by the UMDNJ Foundation, grant no. PC25-11.

Author Contributions M.S. and D.L. cloned mtRNAP variants. M.S., D.L., D.T. and Y.I.M. performed RNAP purification and biochemical assays. R.R. and D.T. prepared the crystals. R.R. performed structure determination and modelling. P.C. and D.T. designed and supervised the project and prepared the manuscript.

Author Information Coordinates and structure factors of the human mtRNAP crystal structure are deposited with the Protein Data Bank under accession code 3SPA. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to P.C. (cramer@lmb.uni-muenchen.de) or D.T. (d.temiakov@umdnj.edu).

Primary motor cortex underlies multi-joint integration for fast feedback control

J. Andrew Pruszynski^{1,2*}, Isaac Kurtzer^{1,3*}, Joseph Y. Nashed¹, Mohsen Omrani¹, Brenda Brouwer^{1,4} & Stephen H. Scott^{1,5,6}

A basic difficulty for the nervous system is integrating locally ambiguous sensory information to form accurate perceptions about the outside world^{1–4}. This local-to-global problem is also fundamental to motor control of the arm, because complex mechanical interactions between shoulder and elbow allow a particular amount of motion at one joint to arise from an infinite combination of shoulder and elbow torques⁵. Here we show, in humans and rhesus monkeys, that a transcortical pathway through primary motor cortex (M1) resolves this ambiguity during fast feedback control. We demonstrate that single M1 neurons of behaving monkeys can integrate shoulder and elbow motion information into motor commands that appropriately counter the underlying torque within about 50 milliseconds of a mechanical perturbation. Moreover, we reveal a causal link between M1 processing and multi-joint integration in humans by showing that shoulder muscle responses occurring ~50 milliseconds after pure elbow displacement can be potentiated by transcranial magnetic stimulation. Taken together, our results show that transcortical processing through M1 permits feedback responses to express a level of sophistication that rivals voluntary control; this provides neurophysiological support for influential theories positing that voluntary movement is generated by the intelligent manipulation of sensory feedback^{6,7}.

Extensive research has shown that some of our fastest motor reactions express a degree of sophistication that rivals voluntary actions^{8–10}, but little is known about the neural substrates that underlie this sophistication¹¹. The present monkey and human studies test the hypothesis that primary motor cortex (M1) provides a neural substrate for integrating shoulder and elbow motion information for fast feedback control, a key ability for generating fast and accurate corrections^{8,12,13}. M1 is a prime candidate to mediate this ability because: (1) it forms part of a transcortical feedback pathway, giving it access to the required afferent information^{14,15}; (2) it is a key node for voluntary control, which appropriately incorporates shoulder and elbow information when generating commands for voluntary actions^{16,17}; and (3) influential theories posit that voluntary movement involves the sophisticated manipulation of sensory information⁶, suggesting substantial functional and anatomical overlap between voluntary and feedback control⁷.

We first determined whether individual neurons in monkey M1 exhibit a pattern of activity consistent with multi-joint motion integration during fast feedback control. Two male rhesus monkeys were trained to counter unpredictable step-torque perturbations applied at the shoulder and/or elbow which displaced their hand from a central target. To receive water reward, the monkeys needed to return their hand to the target within 750 ms and remain within it for an additional 3 s, allowing us to analyse both fast feedback responses (<100 ms post-perturbation) and steady-state motor outputs (last 1 s of stabilization) produced in response to the applied torque within the same trial.

Because our experimental scheme (Fig. 1a) was specifically designed to examine sensorimotor control of the shoulder joint, we were

principally interested in neurons whose steady-state motor outputs varied with the exerted shoulder torque (that is, shoulder-like neurons). As in our previous studies, we found that the population of neurons was biased towards combined shoulder and elbow torques (Rayleigh test for bimodality, $P < 0.05$), making shoulder-like neurons relatively rare¹⁸. In total, 25 of 356 M1 neurons were categorized as shoulder-like because they exhibited significant directional tuning to steady-state loads (plane-fit, $P < 0.05$) and a preferred torque direction within 15° of either shoulder flexion or shoulder extension torque (Supplementary Fig. 1).

The key question is how quickly shoulder-like neurons become selectively tuned to shoulder torque following an unexpected torque perturbation. This is not a mere restatement of our selection criteria

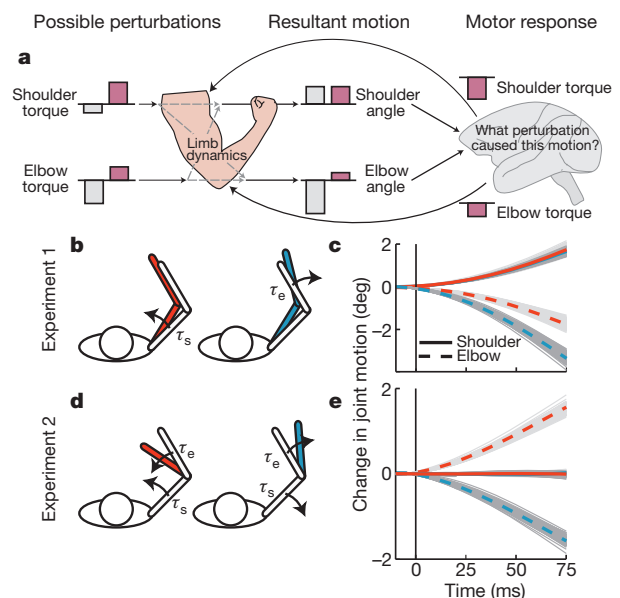


Figure 1 | Experimental methods. **a**, Because of the mechanical properties of the limb, an infinite combination of possible shoulder and elbow torque perturbations can cause the same resultant shoulder motion. Generating the shoulder motor response which counters the applied shoulder torque perturbation requires integrating elbow information. **b**, Limb configuration before (unfilled) and after (filled) a torque perturbation was applied at either the shoulder (τ_s) or elbow (τ_e) as in experiment 1. Opposite conditions (shoulder-extensor / elbow-flexor torque) not shown. **c**, Joint displacement resulting from the shoulder (red) and elbow (blue) perturbation conditions in **b**. The perturbations yielded similar shoulder motion but substantially different elbow motion. Solid lines represent the mean displacements and the grey lines show individual trials. **d**, Limb configuration before and after a multi-joint flexion or multi-joint extension torque perturbation as in experiment 2. **e**, The perturbations shown in **d** caused substantial elbow motion but almost no shoulder motion.

¹Centre for Neuroscience Studies, Queen's University, Kingston, Ontario, Canada K7L 3N6. ²Department of Integrative Medical Biology, Physiology Section, Umeå University, Umeå SE90 187, Sweden. ³Department of Neuroscience and Histology, New York College of Osteopathic Medicine, Old Westbury, New York 11568-8000, USA. ⁴School of Rehabilitation Therapy, Queen's University, Kingston, Ontario, Canada K7L 3N6. ⁵Department of Biomedical and Molecular Sciences, Queen's University, Kingston, Ontario, Canada K7L 3N6. ⁶Department of Medicine, Queen's University, Kingston, Ontario, Canada K7L 3N6.

*These authors contributed equally to this work.

because local shoulder information is sufficient for countering the underlying shoulder torque in the steady-state. In contrast, the only way that fast feedback responses can account for the ambiguous relationship between local joint motion and global torque is by integrating information from both the shoulder and elbow.

The need to resolve ambiguous motion information for fast feedback control is exemplified in our first experiment, where we applied either shoulder torque or elbow torque perturbations (Fig. 1b). These perturbations caused substantially different amounts of elbow motion but nearly identical shoulder motion (Fig. 1c). If shoulder-like neurons integrate both shoulder and elbow motion information, then they should differentiate between the two conditions and respond more strongly to the shoulder torque perturbation than the elbow torque perturbation. Figure 2a presents an exemplar neuron which follows this pattern. It was maximally active during steady-state compensation of shoulder-extension torque, and it responded more strongly to the shoulder-extension torque perturbation than the elbow-flexion torque perturbation (*t*-test, $t_{18} = 2.2$, $P < 0.05$) within ~ 60 ms of perturbation onset. Moreover, the population of shoulder-like neurons (Fig. 2b, Supplementary Fig. 2a) also quickly expressed greater activity for shoulder torque perturbations than elbow torque perturbations (paired *t*-test, $t_{24} = 2.7$, $P < 0.01$; 15 out of 20 neurons in monkey P and 4 out of 5 neurons in monkey X show the expected trend).

Our hypothesis makes the additional prediction that different amounts of inhibition should occur to torque perturbations that are opposite to a neuron's steady-state preference. That is, a neuron which is maximally active during steady-state compensation of shoulder-extension torque should quickly express more inhibition to shoulder-flexion torque perturbations than to elbow-extension torque perturbations. This prediction was verified across the population (paired *t*-test, $t_{24} = 2.1$, $P < 0.05$), demonstrating that shoulder-like neurons possess a pattern of multi-joint integration appropriate for both excitatory and inhibitory perturbations (Supplementary Fig. 2a).

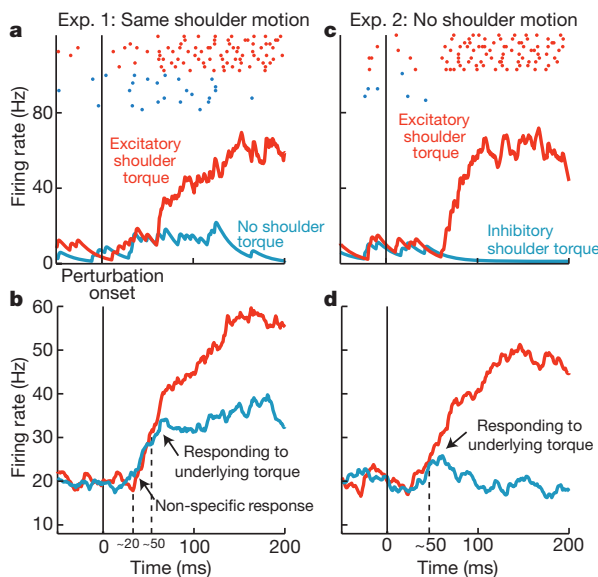


Figure 2 | Neurons in primary motor cortex. **a**, Responses of an exemplar shoulder-like neuron to either an elbow (blue) or shoulder (red) torque perturbation in experiment 1 where shoulder or elbow torque perturbations yielded the same shoulder motion. Data aligned on perturbation onset. Dots represent single action potentials and the trace depicts the average response. **b**, Same format as **a** but representing the population response across shoulder-like neurons. The arrows indicate when neural activity is non-specific to the underlying shoulder torque perturbation and then when it begins to respond appropriately to the underlying shoulder torque. **c**, **d**, Same format as **a** and **b** but for experiment 2 where combined shoulder and elbow torque perturbations yielded no shoulder motion.

Another situation where the nervous system must resolve locally ambiguous information is exemplified in our second experiment, in which torque perturbations at both the shoulder and elbow cause substantial elbow motion but no shoulder motion (Fig. 1d,e). If fast feedback responses of shoulder-like neurons appropriately integrate shoulder and elbow motion to counter the underlying torque then they should respond to this perturbation, even though the shoulder joint is not displaced and no local shoulder sensor (in the muscle, joint or skin) can signal the event. Indeed, the exemplar neuron (Fig. 2c) increased its activity within ~ 60 ms of pure elbow extension motion, which is appropriate for countering the underlying shoulder extensor torque and consistent with its response in our first experiment (*t*-test, $t_{18} = 5.1$, $P < 10^{-3}$). The population of shoulder-like neurons also quickly expressed the predicted response pattern (Fig. 2d, Supplementary Fig. 2b; paired *t*-test, $t_{24} = 4.4$, $P < 10^{-3}$; 15 out of 20 neurons in monkey P and 4 out of 5 neurons in monkey X show the expected trend).

The above analysis established that M1 neurons integrate shoulder and elbow motion to counter the underlying torque perturbation within a binned epoch from 50 to 100 ms after perturbation onset (Fig. 3a). The response of shoulder-like neurons in this epoch paralleled the response of monkey shoulder muscles in the same epoch (Fig. 3b), suggesting that M1 contributes to the observed muscle activity. To provide further evidence of a functional link, we calculated the temporal evolution of multi-joint integration for both neurons and muscles using a receiver operating characteristic (ROC) analysis^{9,19}. We found that multi-joint integration occurred in the population of M1 neurons 8–20 ms before it occurred in muscles (Fig. 3c). Although

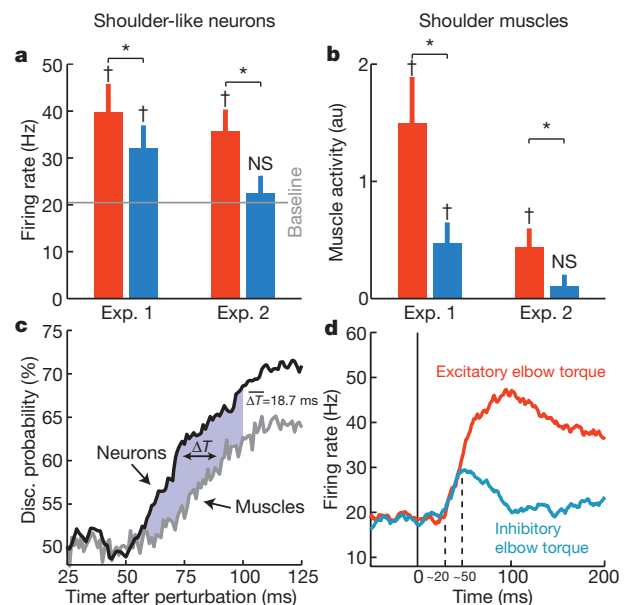


Figure 3 | Population analysis of neurons and muscles. **a**, Binned response (50–100 ms post-perturbation) across shoulder-like neurons. For Experiment 1, the red and blue bars represent responses to shoulder and elbow torque perturbations, respectively. For Experiment 2, the red and blue bars depict responses to pure elbow motion caused by a torque perturbation aligned with or opposite to the neuron's steady-state preference, respectively. Error bars, s.e.m.; *significant differences between conditions (paired *t*-test, $P < 0.05$); †significant differences from baseline; NS, not significant. **b**, Same format as **a** but for the population of muscles. Because of the normalization procedure, muscle baseline activity is 0 arbitrary units (a.u.). **c**, Average discrimination probability over time for the population of neurons and muscles as determined by ROC analysis. Conditions are collapsed across experiments such that the vertical axis is a metric of multi-joint integration. On average, the neurons led the muscles by ~ 19 ms as estimated by the mean temporal difference between the neural and muscle ROC curves (ΔT) from 50 to 100 ms post-perturbation (shaded area). **d**, Same format as Fig. 2d but for elbow-like neurons.

this temporal lead is substantially shorter than typical measurements of M1-to-muscle latency at the initiation of voluntary movement (~ 60 ms)²⁰, it is consistent with the known conduction delay between M1 neurons and muscles of the monkey upper-limb^{15,21}.

Interestingly, shoulder-like M1 neurons did not immediately account for the limb's mechanical properties, displaying a non-specific response to the torque perturbations from ~ 20 to 50 ms following perturbation onset (Fig. 2b, d). A similar non-specific response was observed across our whole population of M1 neurons (Supplementary Fig. 2c,d). For example, neurons which preferentially responded to steady-state elbow torque (elbow-like neurons, Supplementary Fig. 1) initially exhibited the same response whether the perturbation flexed or extended the elbow joint as in our second experiment (Fig. 3d). This non-specific response is strikingly similar to a population of neurons in primary visual cortex (V1), which initially respond ambiguously to objects placed in their receptive field and become sensitive to motion direction only after 20–30 ms (ref. 2), a delay attributed to interactions among V1 neurons²². The temporal evolution of multi-joint integration that we observe may also reflect processing intrinsic to M1 or it may be caused by delayed contributions from other neural structures, such as somatosensory cortex and cerebellum; this is an important issue that warrants further investigation.

Although the activity of single neurons in monkeys provides evidence that M1 is functionally linked to multi-joint integration for fast feedback control, the above data are ultimately correlational and cannot establish whether M1 causes the co-varying pattern of shoulder muscle activity. We addressed this issue by directly influencing the processing of M1 in human participants while they generated feedback corrections similar to those in the monkey study. Applying a single pulse of transcranial magnetic stimulation (TMS) over M1 will excite its intrinsic circuits and, thereby, evoke a burst of muscle activity. When TMS is applied in conjunction with a joint perturbation, the response in the stretched muscle is much larger than the linear sum of the response to TMS alone and the perturbation alone^{23–25}. These supra-linear effects—previously demonstrated in the finger, wrist and elbow muscles—occur only when TMS is timed to evoke a response > 50 ms after the perturbation, suggesting that the two stimuli interact through a common cortical circuit and that feedback control at latencies > 50 ms reflects processing in M1. We established the validity of this technique for shoulder muscles, as a supra-linear response occurred when TMS was delivered ~ 65 ms after the shoulder muscle was stretched (that is, during the long-latency reflex; *t*-test; extensor: $t_9 = 6.7$, $P < 10^{-3}$; flexor: $t_8 = 6.0$, $P < 10^{-3}$) but not 25 ms after the muscle was stretched (that is, during the short-latency reflex; extensor: $t_9 = -0.5$, $P > 0.5$; flexor: $t_8 = 0.5$, $P > 0.5$), when only spinal processes could contribute (Supplementary Fig. 3, left column).

The critical question is whether M1 causally contributes to multi-joint integration for fast feedback control. We tested this hypothesis by applying TMS in conjunction with the torque perturbation that causes pure elbow displacement (Fig. 1d, e). Any supra-linearity of the shoulder muscle response in this condition must reflect afferent information from the elbow joint onto cortical circuits controlling shoulder muscles, because local shoulder afferents are not physically affected by pure elbow motion. The predicted supra-linear effect was observed for both shoulder flexors and extensors (Fig. 4, Supplementary Fig. 3, right column) with TMS delivered 65 ms after perturbation onset (extensor: $t_9 = 3.8$, $P < 0.01$; flexor: $t_8 = 5.3$, $P < 10^{-3}$) but not 25 ms after perturbation onset (extensor: $t_9 = -2.7$, $P > 0.5$; flexor: $t_8 = -0.1$, $P > 0.5$). The observed supra-linearity probably reflects latency-specific engagement of M1 rather than a general change in motor neuron excitability, because we found no correlation between the magnitude of perturbation-evoked activity and the amount of supra-linearity at either latency ($P > 0.1$, Supplementary Fig. 4). Taken together, these results provide strong evidence that M1 causally underlies multi-joint integration for fast feedback control.

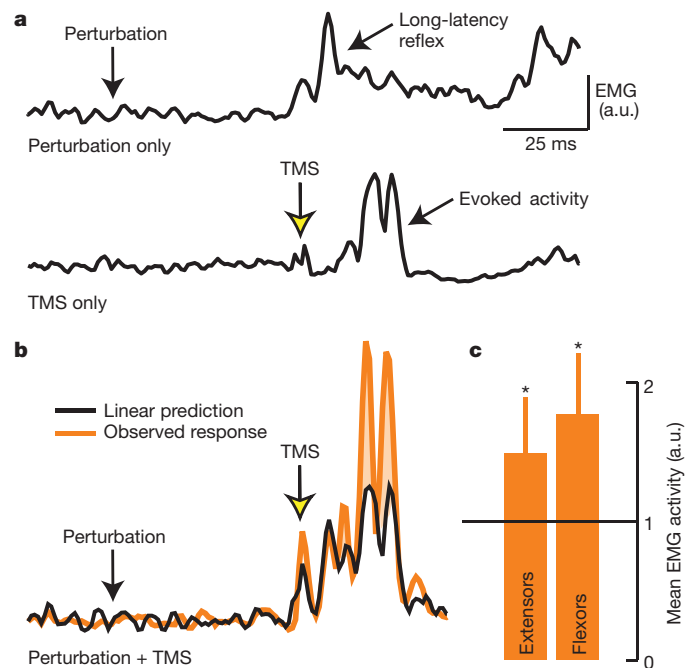


Figure 4 | TMS and perturbation evoked activity in human shoulder muscles. **a**, Response evoked in an exemplar shoulder muscle (posterior deltoid) when a mechanical perturbation or TMS was applied in isolation. Note that the mechanical perturbation yielded a robust long-latency reflex and that the TMS yielded substantial evoked activity. **b**, Observed response (orange) and linear prediction (sum of responses in **a**, black) when the mechanical perturbation and TMS were applied in the same trial. **c**, Group muscle response (mean + s.d.) when TMS was paired with the perturbation normalized by the sum of their separate effects ($E_{\text{norm}} = E_{\text{TMS,pert}} / (E_{\text{TMS}} + E_{\text{pert}})$). Values above 1 indicate supra-linearity; *significant supra-linearity (*t*-test, $P < 0.05$).

Previous studies have demonstrated that fast feedback responses in M1 are scaled by task-constraints, such as movement amplitude²⁶, surface texture²⁷ and intended vigour^{14,28}. Our results show that M1 also integrates locally-ambiguous motion information into a global response that accounts for the limb's mechanical properties, a more complex capability that is central to successfully guiding whole-arm movements¹⁷. It is well established that the voluntary motor system accounts for the mechanical properties of the limb and that this capability is expressed in the activity of M1 neurons²⁹. We have previously argued that the functional similarity of voluntary and feedback control is not an accident and probably arises because of a common neural implementation that includes M1⁷. This expectation is consistent with recent theories of sensorimotor control, which posit that voluntary behaviour involves the sophisticated manipulation of sensory information⁶. If our suggestion is true, then feedback processing in M1 should possess all the capabilities of voluntary processing in M1 and, likewise, studying feedback processing may provide a useful window into voluntary control.

METHODS SUMMARY

Studies were approved by the Queen's University Research Ethics Board and Animal Care Committee. Ten human subjects gave informed consent and completed the experiments as previously described⁸. Two monkeys (*Macaca mulatta*, ~ 10 kg) performed a similar experimental scheme in a miniaturized version of the same apparatus with $\sim 10\times$ smaller loads (KINARM, BKIN Technologies)³⁰. Monkeys did not counter a pre-perturbation background load, and were exposed to eight randomly-interleaved step-torque perturbations, four used in experiment 1 and two used in experiment 2. All eight conditions were used to calculate steady-state tuning by performing a planar regression on the neural activity when the monkey had re-stabilized its hand at the central target¹⁸.

Neural recordings were performed with single electrodes and processed according to standard techniques^{18,29}. Monkey muscle activity was acquired from mono-articular shoulder muscles (anterior/middle/posterior deltoid, pectoralis major;

$n = 34$) using fine-wire electrodes. Human experiments used surface electrodes (posterior deltoid, pectoralis major; $n = 19$). Population responses for both neurons and muscles were collapsed across shoulder flexion and extension conditions according to their predicted excitatory and inhibitory effects (for example, shoulder-flexion torque perturbations were excitatory for shoulder extensor muscles/neurons and inhibitory for shoulder flexor muscles/neurons).

Single pulses of TMS (MES-10, Cadwell) were applied over left M1 with a posterior orientation of 30–45°. Placement/orientation of the double coil was chosen to evoke the largest response from the muscle of interest, ~4.5 cm lateral from vertex. Stimulation magnitude (40–50% of maximum) was selected to deliver the smallest-possible consistent response (evoked activity on seven consecutive trials) when the shoulder muscle countered a steady-state load. TMS only, perturbation only and combined TMS and perturbation trials were randomly interleaved. TMS was timed to evoke shoulder muscle activity either ~25 ms or ~65 ms after perturbation onset.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 7 March; accepted 10 August 2011.

Published online 28 September 2011.

- Green, A. M. & Angelaki, D. E. Multisensory integration: resolving sensory ambiguities to build novel representations. *Curr. Opin. Neurobiol.* **20**, 353–360 (2010).
- Pack, C. C., Livingstone, M. S., Duffy, K. R. & Born, R. T. End-stopping and the aperture problem: two-dimensional motion signals in macaque V1. *Neuron* **39**, 671–680 (2003).
- Pei, Y. C., Hsiao, S. S., Craig, J. C. & Bensmaia, S. J. Neural mechanisms of tactile motion integration in somatosensory cortex. *Neuron* **69**, 536–547 (2011).
- Angelaki, D. E., Shaikh, A. G., Green, A. M. & Dickman, J. D. Neurons compute internal models of the physical laws of motion. *Nature* **430**, 560–564 (2004).
- Hollerbach, J. M. & Flash, T. Dynamic interactions between limb segments during planar arm movement. *Biol. Cybern.* **44**, 67–77 (1982).
- Todorov, E. & Jordan, M. I. Optimal feedback control as a theory of motor coordination. *Nature Neurosci.* **5**, 1226–1235 (2002).
- Scott, S. H. Optimal feedback control and the neural basis of volitional motor control. *Nature Rev. Neurosci.* **5**, 532–546 (2004).
- Kurtzer, I. L., Pruszynski, J. A. & Scott, S. H. Long-latency reflexes of the human arm reflect an internal model of limb dynamics. *Curr. Biol.* **18**, 449–453 (2008).
- Pruszynski, J. A., Kurtzer, I. & Scott, S. H. Rapid motor responses are appropriately tuned to the metrics of a visuospatial task. *J. Neurophysiol.* **100**, 224–238 (2008).
- Cole, K. J., Gracco, V. L. & Abbs, J. H. Autogenic and nonautogenic sensorimotor actions in the control of multiarticulate hand movements. *Exp. Brain Res.* **56**, 582–585 (1984).
- Shemmell, J., Krutky, M. A. & Perreault, E. J. Stretch sensitive reflexes as an adaptive mechanism for maintaining limb stability. *Clin. Neurophysiol.* **121**, 1680–1689 (2010).
- Gielen, C. C., Ramaekers, L. & van Zuylen, E. J. Long-latency stretch reflexes as coordinated functional responses in man. *J. Physiol. (Lond.)* **407**, 275–292 (1988).
- Soechting, J. F. & Lacquaniti, F. Quantitative evaluation of the electromyographic responses to multidirectional load perturbations of the human arm. *J. Neurophysiol.* **59**, 1296–1313 (1988).
- Evarts, E. V. & Tanji, J. Reflex and intended responses in motor cortex pyramidal tract neurons of monkey. *J. Neurophysiol.* **39**, 1069–1080 (1976).
- Cheney, P. D. & Fetz, E. E. Corticomotoneuronal cells contribute to long-latency stretch reflexes in the rhesus monkey. *J. Physiol. (Lond.)* **349**, 249–272 (1984).
- Porter, R. & Lemon, R. N. *Corticospinal Function and Voluntary Movement* (Oxford Univ. Press, 1993).
- Scott, S. H. The role of primary motor cortex in goal-directed movements: insights from neurophysiological studies on non-human primates. *Curr. Opin. Neurobiol.* **13**, 671–677 (2003).
- Herter, T. M., Korb, T. & Scott, S. H. Comparison of neural responses in primary motor cortex to transient and continuous loads during posture. *J. Neurophysiol.* **101**, 150–163 (2009).
- Green, D. M. & Swets, J. A. *Signal Detection Theory and Psychophysics* (Wiley, 1966).
- Thach, W. T. Correlation of neural discharge with pattern and force of muscular activity, joint position, and direction of intended next movement in motor cortex and cerebellum. *J. Neurophysiol.* **41**, 654–676 (1978).
- Fetz, E. E. & Cheney, P. D. Postspike facilitation of forelimb muscle activity by primate corticomotoneuronal cells. *J. Neurophysiol.* **44**, 751–772 (1980).
- Knierim, J. J. & Van Essen, D. C. Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *J. Neurophysiol.* **67**, 961–980 (1992).
- Day, B. L., Riescher, H., Struppler, A., Rothwell, J. C. & Marsden, C. D. Changes in the response to magnetic and electrical stimulation of the motor cortex following muscle stretch in man. *J. Physiol. (Lond.)* **433**, 41–57 (1991).
- Palmer, E. & Ashby, P. Evidence that a long latency stretch reflex in humans is transcortical. *J. Physiol. (Lond.)* **449**, 429–440 (1992).
- Lewis, G. N., Polych, M. A. & Byblow, W. D. Proposed cortical and sub-cortical contributions to the long-latency stretch reflex in the forearm. *Exp. Brain Res.* **156**, 72–79 (2004).
- Fromm, C. & Evarts, E. V. Relation of motor cortex neurons to precisely controlled and ballistic movements. *Neurosci. Lett.* **5**, 259–265 (1977).
- Picard, N. & Smith, A. M. Primary motor cortical responses to perturbations of prehension in the monkey. *J. Neurophysiol.* **68**, 1882–1894 (1992).
- Evarts, E. V. Motor cortex reflexes associated with learned movement. *Science* **179**, 501–503 (1973).
- Scott, S. H., Gribble, P. L., Graham, K. M. & Cabel, D. W. Dissociation between hand motion and population vectors from neural activity in motor cortex. *Nature* **413**, 161–165 (2001).
- Scott, S. H. Apparatus for measuring and perturbing shoulder and elbow joint positions and torques during reaching. *J. Neurosci. Methods* **89**, 119–127 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the Canadian Institutes of Health Research (CIHR) and the National Sciences and Engineering Research Council of Canada (NSERC). J.A.P., I.K. and M.O. received salary awards from CIHR. We thank K. Moore and J. Peterson for technical support, and G. Blohm, B. Edin, R. Flanagan, R. Johansson, A. Kahn and D. Munoz for comments on the manuscript.

Author Contributions J.A.P. collected and analysed the monkey data, helped design the experiments, interpreted the results and helped write the manuscript. I.K. collected and analysed the human data, helped design the experiments, interpreted the results and helped write the manuscript. J.Y.N. helped collect the human data. M.O. helped collect a subset of the monkey data. B.B. helped interpret the human data. S.H.S. helped design the experiments, interpreted the results and helped write the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests; details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.H.S. (steve.scott@queensu.ca).

METHODS

Participants and apparatus. The studies presented were approved by the Queen's University Research Ethics Board. All monkey ($n = 2$, *Macaca mulatta*, ~10 kg, male) procedures were approved by the Queen's University Animal Care Committee. Human subjects (6 females, 4 males, median age = 27) were neurologically unimpaired, had normal/corrected vision and provided informed consent. Human and monkey experiments were performed using different versions of the same robotic exoskeleton (KINARM, BKIN Technologies) which allows combined flexion and extension movements of the shoulder and elbow in the horizontal plane and can independently apply mechanical loads to the shoulder and/or elbow³⁰. Target lights and simulated hand feedback were presented in the horizontal plane of the task via a virtual reality display, and direct vision of the hand was limited either by a physical barrier (humans) or by a lack of ambient light (monkeys).

Neural, muscle and kinematic recordings. Recording chambers were surgically implanted under inhalation anaesthetic, and neural recordings were performed according to standard techniques^{18,29}. Single tungsten microelectrodes (FHC) were advanced until neural activity was observed. Individual neurons were then isolated and neural activity was recorded from those neurons with clear responses to either passive or active movements of the shoulder and/or elbow. Neurons which primarily responded to motion of the wrist or fingers were not recorded. Neurons recorded in the task ($n = 356$) were located in the rostral bank of the central sulcus as well as more superficial sites where previous mapping efforts showed that trains of electrical stimulation (11 pulses, 333 Hz, 0.2 ms pulse width, <50 μ A) could elicit shoulder and/or elbow movement. Post-mortem histology confirmed that recording sites from monkey P were located in M1.

Monkey muscle activity was acquired from mono-articular shoulder muscles (anterior deltoid, middle deltoid, posterior deltoid, pectoralis major; $n = 34$) using fine-wire electrodes. Electrodes consisted of two single-strand wires and were individually inserted into the muscle belly spaced ~5 mm apart. Insertion was guided by anatomical landmarks and was confirmed by microstimulation. Human experiments used surface electrodes (Bortec AMT-8) and focused on those mono-articular muscles which could be readily recorded from the surface (posterior deltoid, pectoralis major; $n = 19$). Muscle activity was recorded at either 4 kHz (monkey) or 1 kHz (human), aligned on perturbation onset and full-wave rectified before analysis⁹. Only those muscles with clear phasic responses to the mechanical perturbation were analysed. Kinematic data and applied torques were acquired directly from the KINARM device and were sampled at the same rate as the muscle activity.

Transcranial magnetic stimulation. We followed standard procedures in the TMS portion of the study^{23–25}. Single pulses of TMS (MES-10, Cadwell) were applied over left M1 with a posterior orientation of 30–45°. Placement and orientation of the double cone coil were chosen to evoke the largest response from the muscle of interest, ~4.5 cm lateral from vertex. Stimulation magnitude was selected to deliver the smallest-possible consistent response (evoked response on seven consecutive stimulations, average of 40% and 51% of the stimulator's maximum output for the posterior deltoid and pectoral major, respectively) when the muscle

of interest actively countered a 3 N m background load (that is, active motor threshold).

Experimental scheme. The experimental procedure and logic have been previously described⁸. The major difference in the human portion of this study was the parallel implementation of TMS. Briefly, subjects stabilized their hand in a small central target while countering a steady state shoulder torque (3 N m) which activated either the shoulder flexor or extensor muscles. After a random hold time (1–4 s), an unpredictable torque pulse (100 ms duration) was introduced and the trial ended when the subjects re-stabilized in the target for 500 ms.

In total, four torque perturbations were used in the human study. Two single-joint torque perturbations (3 N m shoulder-flexion for shoulder flexor muscles and 3 N m shoulder-extension for shoulder extensor muscles) made up experiment 1, and two multi-joint torque perturbations (3 N m shoulder-flexion/3 N m elbow-flexion for shoulder flexor muscles; 3 N m shoulder-extension/3 N m elbow-extension for shoulder extensor muscles) made up experiment 2. Perturbation only, TMS only and combined TMS and perturbation trials were randomly interleaved. In combined TMS and perturbation trials, the TMS was timed to evoke shoulder muscle activity either ~25 ms or ~65 ms after perturbation onset.

The perturbation, TMS placement and TMS intensity were chosen for the shoulder flexor muscle and shoulder extensor muscle in two successive blocks. Half the subjects began with the conditions for the shoulder flexor muscle and half the subjects began with the conditions for the shoulder extensor muscle. Thirty repeats of the 14 conditions were collected for a total of 420 trials in a session that lasted about 2.5 h.

Monkeys performed a similar scheme with ~10 \times smaller loads. Unlike humans, the monkeys did not counter a pre-perturbation background load and they were exposed to eight randomly-interleaved step-torque perturbations ([shoulder torque, elbow torque], applied flexion/extension = positive/negative: 1. [0.28 N m, 0 N m], 2. [0.24, 0.24], 3. [0, 0.24], 4. [−0.2, 0.2], 5. [−0.28, 0], 6. [−0.24, −0.24], 7. [0, −0.24], 8. [0.2, −0.2]) and catch trials where no perturbations occurred. Four of these perturbations (1,3,5,7) formed experiment 1 and two (2,6) formed experiment 2. All eight conditions were used to calculate the steady-state tuning of each neuron by performing a planar regression on the neural activity when the monkey had re-stabilized its hand at the central target¹⁸. To receive water reward, the monkeys needed to return their hand to the target within 750 ms and remain within it for an additional 3 s, allowing us to analyse both fast feedback responses (<100 ms post-perturbation) and steady-state motor outputs (last 1 s of stabilization) to the applied torque within the same trial. Five to twenty repeats were collected per experimental condition.

Population responses for both muscles and neurons were calculated by collapsing across shoulder flexion and extension conditions according to their predicted excitatory and inhibitory effects. That is, applied shoulder-flexion torque perturbations were excitatory for shoulder extensor muscles/neurons and inhibitory for shoulder flexor muscles/neurons. Applied shoulder-extension torque perturbations were excitatory for shoulder flexor muscles/neurons and inhibitory for shoulder extensor muscles/neurons.

Theta-paced flickering between place-cell maps in the hippocampus

Karel Jezek^{1,2}, Espen J. Henriksen¹, Alessandro Treves^{1,3}, Edvard I. Moser¹ & May-Britt Moser¹

The ability to recall discrete memories is thought to depend on the formation of attractor states in recurrent neural networks^{1–4}. In such networks, representations can be reactivated reliably from subsets of the cues that were present when the memory was encoded, at the same time as interference from competing representations is minimized. Theoretical studies have pointed to the recurrent CA3 system of the hippocampus as a possible attractor network^{3,4}. Consistent with predictions from these studies, experiments have shown that place representations in CA3 and downstream CA1 tolerate small changes in the configuration of the environment but switch to uncorrelated representations when dissimilarities become larger^{5–9}. However, the kinetics supporting such network transitions, at the subsecond timescale, is poorly understood. Here we show in rats that instantaneous transformation of the spatial context does not change the hippocampal representation all at once but is followed by temporary bistability in the discharge activity of CA3 ensembles. Rather than sliding through a continuum of intermediate activity states, the CA3 network undergoes a short period of competitive flickering between preformed representations of the past and present environment before settling on the latter. Network flickers are extremely fast, often with complete replacement of the active ensemble from one theta cycle to the next. Within individual cycles, segregation is stronger towards the end, when firing starts to decline, pointing to the theta cycle as a temporal unit for expression of attractor states in the hippocampus. Repetition of pattern-completion processes across successive theta cycles may facilitate error correction and enhance discriminative power in the presence of weak and ambiguous input cues.

The place-cell population of the hippocampus is thought to create a neural representation of the spatial environment¹⁰. Accumulating evidence indicates that environments are generally represented in hippocampal cells by many discrete maps, each corresponding to a distinct environment or a unique experience within the environment^{5,9,11–13}. Which map is active at any given time depends on external sensory inputs as well as recent history^{7,14}. Incongruity between the active map and sensory inputs may lead to partial or complete replacement of the active representation^{15,16}. The kinetics of map substitutions has remained elusive owing to a shortage of experimental and analytical tools for subsecond-timescale neural population analyses.

Here, we developed tools to determine how local network activity evolves in the hippocampus in response to sudden changes in the cues that define spatial context. Rats with tetrodes in CA3 (Supplementary Fig. 1) were first trained on separate trials in two boxes with different sets of light cues in a dark room (boxes A and B). The procedure favours the development of uncorrelated place representations in A and B^{5,9} (Supplementary Fig. 2). After several days of training in each box, the rats started in one of the environments; then, after 40–60 s, the cues were switched instantaneously to those of the other environment, effectively ‘teleporting’ the rat from A to B or vice versa. A total of 169 such trials, hereafter referred to as teleportation trials, were performed,

with an average of 33 ± 3 active cells per day (mean \pm s.e.m.; total of 358 active cells; 11 days; 6 rats).

To examine the evolution of network activity after the cue change, we first established expected firing patterns for all locations in each environment. Firing rates were determined for each cell in each spatial bin of each box on separate reference trials (30×30 bins; Fig. 1a). As expected for CA3 place cells¹⁷, the subsets of active cells in the two boxes overlapped minimally: cells with activity in both A and B generally fired at unrelated positions (spatial correlation: 0.112 ± 0.019). The nearly orthogonal nature of the baseline representations made it possible, in principle, to infer from any subsequent cell sample which of the two environments was represented in the hippocampal network at the time of recording. Thus, in the next step, we compared the evolution of activity over successive theta cycles in the teleportation test with activity at corresponding locations in the reference environments. The theta phase with the lowest overall firing rate was used to segment the recording from each teleportation trial into individual theta cycles (Fig. 1b, c and Supplementary Fig. 3). A population vector was then constructed for each theta cycle, consisting of the number of spikes of each of the C simultaneously recorded cells (Fig. 1c). For each theta cycle, the population vector was correlated with vectors of average firing rates (in Hz) for the same location in each reference recording. We first used a simple linear measure of correlation, the un-normalized dot product (DP), between the test population vector \mathbf{r} and each of the two reference vectors, \mathbf{r}_A and \mathbf{r}_B . If \mathbf{r}_A and \mathbf{r}_B were strictly orthogonal and after teleportation test vectors were to be linear combinations $\mathbf{r} \approx c_A \mathbf{r}_A + c_B \mathbf{r}_B$, the DP measure would allow a direct estimate of the coefficients $c_A(t)$ and $c_B(t)$, that is, of the strength with which the two representations are relayed to downstream neurons at time t ^{2,4}. Only trials with continuous theta activity were analysed (149 out of 169 teleportation trials).

The switching of spatial cues was followed by a characteristic pattern of network activity (Fig. 2). Before the transformation, most theta cycles correlated strongly with reference vectors for the first environment (denoted as I, whether A or B) but not for the second (II), as expected (Figs 1d and 2a, b). After cue switching, most theta cycles correlated with the reference vectors for II but the magnitude of the dot product was variable and the network occasionally ‘flickered’ back to strong correlations with I. To investigate the statistics of this flickering, we plotted for each theta cycle the correlation of the momentary population vector with the reference vectors from the same environment (x -axis) and the alternative environment (y -axis) (Fig. 2c). The analysis confirmed that cue switching increased the proportion of cycles correlated with the alternative environment but also identified a subset of cycles that correlated modestly with both environments at the same time. This led us to ask if the instantaneous activity during the transition reflected a simple linear combination of fragments of the A and B representations, as would be expected in the absence of population coherence. We consequently determined whether, in the pooled data, the proportion of theta cycles correlating with both environments at

¹Kavli Institute for Systems Neuroscience and Centre for the Biology of Memory, Norwegian University of Science and Technology, Olav Kyrres gate 9, MTF5, 7489 Trondheim, Norway. ²Department of Physiology of Memory, Institute of Physiology, Videnska 1083, Academy of Sciences of the Czech Republic, 142 20 Prague 4-Krc, Czech Republic. ³Cognitive Neuroscience Sector, SISSA International School for Advanced Studies, via Bonomea 265, 34136 Trieste, Italy.

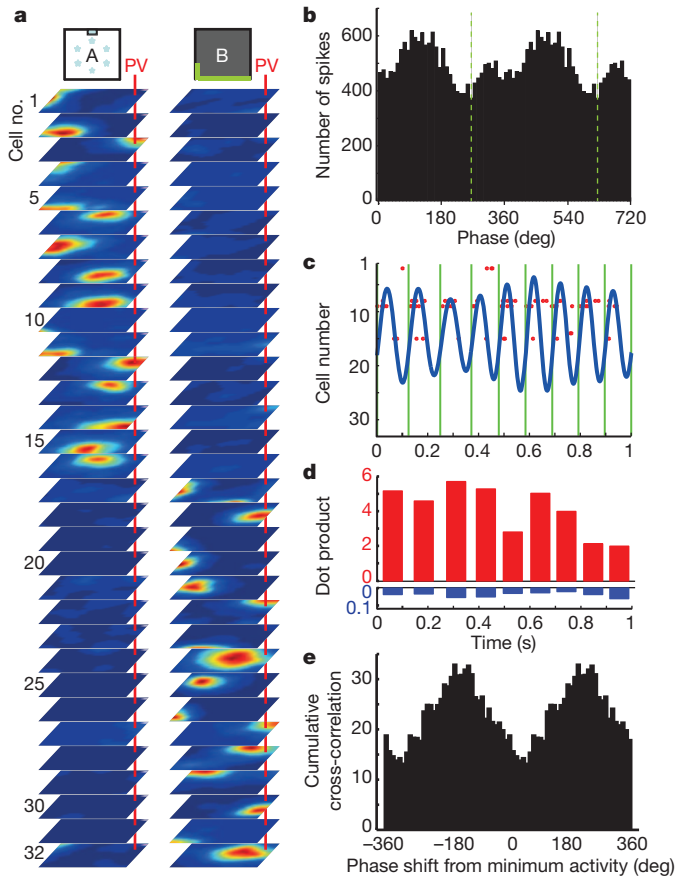


Figure 1 | Procedures for analysing hippocampal transition dynamics.

a, Stack of firing-rate maps in box A (white floor lights; left) and box B (green wall lights; right) for an example set of 32 simultaneously recorded hippocampal CA3 units. Each map shows a colour-coded distribution of firing rates across the square test box (blue, silent; red, maximum). Red line, one of 30×30 population vectors (PV) constructed from the activity of the entire cell ensemble in a given 2×2 cm position bin. Note strong difference in population vectors for A and B.

b, Theta phase modulation for all pyramidal cells on a representative trial. Spike number is shown as a function of theta phase (bin size 10 deg). Dashed green line, phase with lowest firing rate, used to define boundary between cycles.

c, Representative spike distribution across theta cycles in the stable state. Rasters of red dots show spike times of individual cells in relation to 6–11 Hz filtered local electroencephalogram (EEG; blue). Green lines indicate theta-cycle boundaries (b). The ensemble distribution of activity during one cycle represents the momentary population vector.

d, Dot-product correlation between momentary population vector and reference vectors at the corresponding position in A (red) and B (blue) during a baseline trial in A.

e, Cumulative product between correlations with each of the reference environments as a function of the phase for segmentation of theta cycles (0, phase of minimum activity).

the same time ($\mathbf{r} \cdot \mathbf{r}_A > C$, $\mathbf{r} \cdot \mathbf{r}_B > C$, referred to as ‘mixed theta cycles’) was lower than expected if each single unit expressed either one or the other representation independently of the other units. A total of 1.25% of the theta cycles in the recorded data were mixed. This number was lower than in 970 out of 1,000 randomly recombined (shuffled) population vectors (that is, $P < 0.03$; Fig. 2d). The separation between the A- and B-correlated representations was strongest when the cycles were chunked at the point of the lowest average firing rate in the population (Fig. 1b, c, e and Supplementary Fig. 3). Thus, mixed representations existed but were rare. Transitions between orthogonal maps tend to occur in an all-or-none manner, with the entire network flickering coherently at time scales of approximately a tenth of a second.

We subsequently examined the evolution of network activity within the theta cycle. Each cycle was divided into two halves, and mixed states were defined as those half-cycles for which $\mathbf{r} \cdot \mathbf{r}_A > C/2$, $\mathbf{r} \cdot \mathbf{r}_B > C/2$. Mixed population vectors were less abundant than

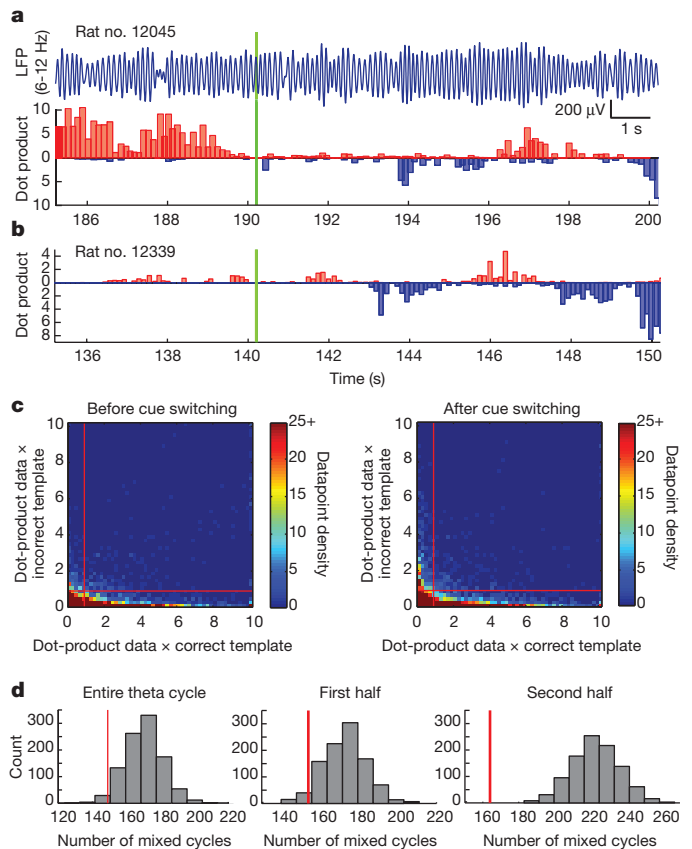


Figure 2 | Theta cycles correlate with either of the reference environments but rarely with both simultaneously.

a, Top, local hippocampal EEG during spatial cue switching from A to B (filtered at 6–11 Hz). Bottom, dot-product correlation between momentary population vectors and reference vectors from A (red) and B (blue) for successive theta cycles before and after cue switching. Dot products are un-normalized (just divided by the number C of recorded cells). All correlations are positive but, for clarity, A and B correlations are plotted in opposite directions. Green line indicates light switch. EEG and ensemble activity were sampled simultaneously. Note that ensemble activity flickered back to the A representation several times after cue switching. Note also the variation in the dot product.

b, Another example of network flickering induced by switching of spatial cues.

c, Matrices reporting the number of cycles falling in each 0.2×0.2 bin of the dot-product correlations $\mathbf{r} \cdot \mathbf{r}_{A/B}$ between momentary population vectors and reference vectors for the present environment (x-axis) or the alternative environment (y-axis). Left, before cue switching. Right, after cue switching (starting from the first cycle correlated with new environment). Note that mixed cycles, defined as cycles with both $\mathbf{r} \cdot \mathbf{r}_A$ and $\mathbf{r} \cdot \mathbf{r}_B$ exceeding C ($x > 1$, $y > 1$; indicated by red lines), were rare.

d, Histograms showing that the number of mixed states after cue switching (red line) is lower than expected from shuffled versions of the same data ($n = 1,000$; grey histogram). Note that mixed states became less frequent during the second half of the theta cycle.

expected from shuffled data during both half-cycles; however, in the shuffled data, the frequency of recombinations with more mixed population vectors than in the observed data increased from 958/1,000 during the first half ($P < 0.05$) to 1,000/1,000 during the second ($P < 0.001$) (Fig. 2d). The low incidence of mixed population vectors at the end of the theta cycle suggests that representations evolve from partially segregated to fully segregated within each activity period.

We then asked how A-correlated and B-correlated theta cycles were organized in time. Because dot products can vary from 0 to indefinitely large, we switched to Pearson product-moment correlations, which by normalizing the correlations to within a fixed $[-1, +1]$ range allow successive theta cycles to be compared more directly (Fig. 3a–c and Supplementary Figs 4–6). Individual theta cycles were now only considered if at least two cells were active (for higher thresholds, see Supplementary Fig. 7). As observed with the dot products, the

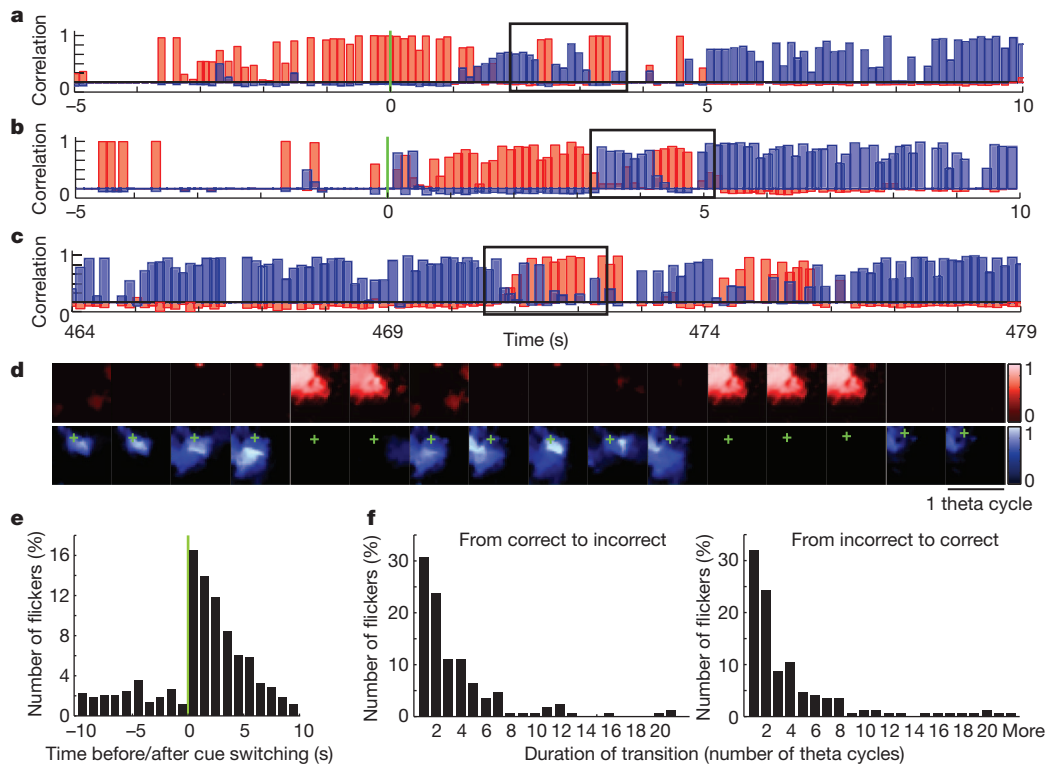


Figure 3 | Temporal dynamics of network flickering. **a–c**, Pearson product-moment correlations showing evolution of population vector correlations after spatial cue switching from box I to II (red, correlation with I; blue, with II; green line, cue switch). Note frequent flickers to the original representation after cue switching in **a** and **b**. **c**, Spontaneous flickering between cue-switch events. Frames: sequences detailed in **d** and Supplementary Fig. 8. **d**, Spatial distribution of correlations between momentary population vectors and reference vectors in the framed area in **a**. Each row shows correlation matrices

for 16 consecutive theta cycles. Top, correlation with I; bottom, with II.

Correlation is colour coded (scale bar). + indicates rat position. **e**, Percentage of flickers to the alternative representation as a function of time before and after cue switching. **f**, Distribution plots showing the time for the network to switch from present ('correct') to past ('incorrect') representation, or vice versa, in number of theta cycles. One cycle means that cycles with alternative representations were consecutive. Note the predominance of immediate transitions.

momentary population vectors correlated strongly with either A or B but rarely with both. Before cue switching, theta cycles were nearly exclusively correlated with reference vectors for the initial environment (I). After cue switching, the network switched almost instantaneously to high correlation with reference vectors for II but then relapsed to I several times during the subsequent seconds (Fig. 3a, b and Supplementary Figs 5 and 6) and occasionally tens of seconds after cue switching (Fig. 3c). These relapses, or flickers, were confined to discrete periods of one or several theta cycles. When the population vector was correlated with reference vectors from other locations in the represented environment, the correlation generally decreased with distance from the animal (Fig. 3d and Supplementary Fig. 8).

To quantify the frequency, timing and duration of flicker events, we defined individual theta cycles as A-correlated if the correlation with reference environment A was above the 95th percentile for $B \times A$ correlations in the reference sessions (that is, more similar to A than 95% of theta cycles in B) and if the correlation with reference environment B was simultaneously below the 5th percentile for $B \times B$ (that is, different from most theta cycles in B) (Supplementary Figs 9 and 10). The analysis showed a clear increase in the frequency of network flickers during the first seconds following the cue switch from I to II, after the network had switched to the II representation for the first time (generally 0–1 s after the cue change; Supplementary Fig. 11 and Supplementary Table 1). The fraction of theta cycles participating in flicker episodes, estimated with Pearson correlations, increased from a stable baseline of 1–3% before cue switching to a level of 10–15% during the first 5 s after the first network switch (Fig. 3e and Supplementary Fig. 12), confirming the tendency revealed by the dot-product analyses (Fig. 2c). The interquartile range of flicker durations increased from 1–1 theta cycles before cue switching to 1–4 during the first 10 s after cue switching (Wilcoxon rank-sum test:

$Z = 2.27$, $P < 0.03$; Supplementary Fig. 11b and Supplementary Table 1). Flicker events were distributed across the entire recording box (Supplementary Fig. 13) and showed no preference for running or heading direction (Supplementary Fig. 14). Flickering was apparent also in CA1 but discrete relapses were clearly less frequent, possibly due to the less reverberating architecture of this subfield (Supplementary Fig. 15).

Transitions between representations occurred within less than a single theta cycle. After the last cycle in a series of non-flicker theta cycles, the alternative representation was fully expressed already in the subsequent theta period in 30.8% of the flicker events where flicker duration could be determined (Fig. 3f). The corresponding percentage of single-cycle returns to the 'correct' representation was 32.0%. In 57.0% of transitions from 'correct' to 'incorrect' reference frame and 62.2% of transitions back from 'incorrect' to 'correct', the time course remained undetermined because the intervening theta cycle contained fewer than two spikes. In each instance of an immediate transition, the network representation was fully developed from the outset, that is, the correlation with the new environment did not increase further within the flicker period ($r = -0.06 \pm 0.11$).

To determine further if the flicker episodes were patterned by the theta oscillation, we finally compared the transition dynamics of our theta-based segmentation procedure with segmentations based on fixed time bins of different width (range 44–500 ms; Supplementary Fig. 3). Among the fixed bins, direct transitions were most abundant when the bins matched the average duration of theta cycles (125 versus 120.4 ms, respectively). The abundance of sharp transitions increased further when the trial was segmented by actual rather than fixed theta periods, with cycles split at the phase with the minimum firing rate (Fig. 1b, c, e and Supplementary Fig. 3). Collectively, these observations suggest strongly that the transitions were paced by the theta rhythm.

Our study provides evidence for competitive interactions between hippocampal representations during changes in spatial reference frame. Although a small subset of the population vectors correlated with both reference environments after the cue change, the number of such mixed states was lower than expected from a sample of independent single units, especially during the second half of each theta cycle. In most cases, the network either switched all at once or flickered between mutually exclusive representations until, after a few seconds, it settled in one of the alternatives. The sustained separation of the neural activity pattern is consistent with the notion that spatial environments, as a whole, are stored as discrete attractors in neural networks of the hippocampus or associated areas such as the entorhinal cortex^{1–4,8,9}. These discontinuities differ from the continuous or quasi-continuous nature of spatial maps for individual environments^{12,18–20}, where sweeps can follow unbroken trajectories, even when retrieval occurs in the absence of actual movement²¹. Changes in attention or experience are likely to generate continuous transitions of the latter type more or less constantly in all environments^{7,22–26}. The low frequency of flickering in the baseline state of the teleportation task suggests, however, that switches to uncorrelated attractor maps are rare and occur primarily when cues are ambiguous or in conflict across sensory modalities.

The time course of flickering episodes has implications for the mechanisms of ensemble activation. It took often only a single theta cycle to fully reactivate a pre-established representation during a flicker event, consistent with models of theta phase precession in which ensembles are activated by propagation through recurrent collaterals after afferent input to a subset of cells early in the theta cycle²⁷. The idea that sensory influences can override attractors at the beginning of the theta cycle, whereas subsequent activity is determined more exclusively by propagation through associative connections, receives further support from the fact that the small number of mixed population vectors occurred primarily during the first half of the theta cycle. The recreation of spatial representations on successive theta cycles in the hippocampus is fundamentally different from the pattern-completion dynamics observed, for example, in the inferior temporal cortex, where the activation may proceed only once per stimulus presentation, at a timescale short enough to facilitate perception but with little opportunity for error correction²⁸. In the hippocampus, repeated convergence to an attractor state might allow the system to self-correct and thereby enhance its discriminative power under conditions where input cues are weak and ambiguous. The present data point to theta cycles as organizational units for this repetitive process but do not preclude additional structure at faster time scales, for example with gamma cycles as units for cell assembly sequences within the theta cycle^{29,30}.

METHODS SUMMARY

Neuronal ensemble activity was recorded from ensembles of CA3 cells while rats foraged in either of two distinct enclosures in a dark room. The enclosures were identical except for internal lights. Testing began by placing the rat in one of the two boxes. After 40–60 s, the light cues were switched, effectively ‘teleporting’ the animal to the other environment. To examine the evolution of network activity, population vectors were defined from the firing rates of all simultaneously recorded cells for every theta cycle before and after the switching of spatial cues. Each population vector was correlated with reference vectors defined from the activity of the same cells at the same spatial location in each box on separate baseline trials.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 25 November 2010; accepted 10 August 2011.

Published online 28 September 2011.

1. Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl Acad. Sci. USA* **79**, 2554–2558 (1982).
2. Amit, D. J., Gutfreund, H. & Sompolinsky, H. Storing infinite numbers of patterns in a spin-glass model of neural networks. *Phys. Rev. Lett.* **55**, 1530–1533 (1985).
3. McNaughton, B. L. & Morris, R. G. M. Hippocampal synaptic enhancement and information storage within a distributed memory system. *Trends Neurosci.* **10**, 408–415 (1987).

4. Treves, A. & Rolls, E. T. Computational constraints suggest the need for two distinct input systems to the hippocampal CA3 network. *Hippocampus* **2**, 189–199 (1992).
5. Muller, R. U. & Kubie, J. L. The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. *J. Neurosci.* **7**, 1951–1968 (1987).
6. Lee, I., Yoganarasimha, D., Rao, G. & Knierim, J. J. Comparison of population coherence of place cells in hippocampal subfields CA1 and CA3. *Nature* **430**, 456–459 (2004).
7. Leutgeb, J. K. *et al.* Progressive transformation of hippocampal neuronal representations in ‘morphed’ environments. *Neuron* **48**, 345–358 (2005).
8. Wills, T. J., Lever, C., Cacucci, F., Burgess, N. & O’Keefe, J. Attractor dynamics in the hippocampal representation of the local environment. *Science* **308**, 873–876 (2005).
9. Colgin, L. L. *et al.* Attractor-map versus autoassociation based attractor dynamics in the hippocampal network. *J. Neurophysiol.* **104**, 35–50 (2010).
10. O’Keefe, J. & Nadel, L. *The Hippocampus as a Cognitive Map* (Oxford Univ. Press, 1978).
11. Gothard, K. M., Skaggs, W. E., Moore, K. M. & McNaughton, B. L. Binding of hippocampal CA1 neural activity to multiple reference frames in a landmark-based navigation task. *J. Neurosci.* **16**, 823–835 (1996).
12. Samsonovich, A. & McNaughton, B. L. Path integration and cognitive mapping in a continuous attractor neural network model. *J. Neurosci.* **17**, 5900–5920 (1997).
13. Derdikman, D. *et al.* Fragmentation of grid cell maps in a multicompartment environment. *Nature Neurosci.* **12**, 1325–1332 (2009).
14. O’Keefe, J. & Speakman, A. Single unit activity in the rat during a spatial memory task. *Exp. Brain Res.* **68**, 1–27 (1987).
15. Gothard, K. M., Skaggs, W. E. & McNaughton, B. L. Dynamics of mismatch correction in the hippocampal ensemble code for space: interaction between path integration and environmental cues. *J. Neurosci.* **16**, 8027–8040 (1996).
16. Skaggs, W. E. & McNaughton, B. L. Spatial firing properties of hippocampal CA1 populations in an environment containing two visually identical regions. *J. Neurosci.* **18**, 8455–8466 (1998).
17. Leutgeb, S., Leutgeb, J. K., Treves, A., Moser, M.-B. & Moser, E. I. Distinct ensemble codes in hippocampal areas CA3 and CA1. *Science* **305**, 1295–1298 (2004).
18. Tsodyks, M. & Sejnowski, T. Associative memory and hippocampal place cells. *Int. J. Neural Syst.* **6**, 81–86 (1995).
19. Romani, S. & Tsodyks, M. Continuous attractors with morphed/correlated maps. *PLoS Comput. Biol.* **6**, e1000869 (2010).
20. McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I. & Moser, M. B. Path integration and the neural basis of the ‘cognitive map’. *Nature Rev. Neurosci.* **7**, 663–678 (2006).
21. Johnson, A. & Redish, A. D. Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *J. Neurosci.* **27**, 12176–12189 (2007).
22. Fenton, A. A. & Muller, R. U. Place cell discharge is extremely variable during individual passes of the rat through the firing field. *Proc. Natl Acad. Sci. USA* **95**, 3182–3187 (1998).
23. Olypher, A. V., Lánský, P. & Fenton, A. A. Properties of the extra-positional signal in hippocampal place cell discharge derived from the overdispersion in location-specific firing. *Neuroscience* **111**, 553–566 (2002).
24. Jackson, J. & Redish, A. D. Network dynamics of hippocampal cell-assemblies resemble multiple spatial maps within single tasks. *Hippocampus* **17**, 1209–1229 (2007).
25. Kelemen, E. & Fenton, A. A. Dynamic grouping of hippocampal neural activity during cognitive control of two spatial frames. *PLoS Biol.* **8**, e1000403 (2010).
26. Blumenfeld, B., Preminger, S., Sagi, D. & Tsodyks, M. Dynamics of memory representations in networks with novelty-facilitated synaptic plasticity. *Neuron* **52**, 383–394 (2006).
27. Tsodyks, M. V., Skaggs, W. E., Sejnowski, T. J. & McNaughton, B. L. Population dynamics and theta rhythm phase precession of hippocampal place cell firing: a spiking neuron model. *Hippocampus* **6**, 271–280 (1996).
28. Akrami, A., Liu, Y., Treves, A. & Jagadeesh, B. Converging neuronal activity in inferior temporal cortex during the classification of morphed stimuli. *Cereb. Cortex* **19**, 760–776 (2009).
29. Harris, K. D., Csicsvari, J., Hirase, H., Dragoi, G. & Buzsáki, G. Organization of cell assemblies in the hippocampus. *Nature* **424**, 552–556 (2003).
30. Colgin, L. L. *et al.* Frequency of gamma oscillations routes flow of information in the hippocampus. *Nature* **462**, 353–357 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Skjerpeng, A. M. Amundsgård, K. Haugen, K. Jenssen, E. Kråkvik and H. Waade for technical assistance. The work was supported by the 7th Framework Programme of the European Commission (‘SPACEBRAIN’, grant agreement no. 200873), an Advanced Investigator Grant to M.-B.M. from the European Research Council (‘ENSEMBLE’), the Kavli Foundation, a Centre of Excellence grant from the Norwegian Research Council, and research projects MSMT CR LC554, 1M0517 and AV0Z50110509 at the Academy of Sciences of the Czech Republic.

Author Contributions K.J., A.T., M.-B.M. and E.I.M. designed the study and discussed analyses and results; K.J. built the apparatus, K.J. and E.J.H. performed experiments, K.J. performed analyses; E.I.M. wrote the paper with input from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to K.J. (karel.jezek@biomed.cas.cz) or E.I.M. (edvard.moser@ntnu.no).

METHODS

Subjects. Six male Long Evans rats (400–500 g at implantation) were housed individually in transparent Plexiglass cages (45 cm × 30 cm × 35 cm). The animals were kept at ~90% of their initial free-feeding body weight and maintained on a 12-h light/12-h dark schedule. All testing occurred in the dark phase. The experiments were performed in accordance with the Norwegian Animal Welfare Act and the European Convention for the Protection of Vertebrate Animals used for Experimental and Other Scientific Purposes.

Electrode preparation and surgery. Neuronal ensemble activity was recorded from ensembles of CA3 or CA1 cells in rats implanted with a 'hyperdrive' containing 14 independently movable tetrodes assembled in a circular bundle. Tetrodes were twisted from four 17- μ m polyimide-coated platinum-iridium wires (90% and 10%, respectively; California Fine Wire Company). Electrode tips were plated with platinum to reduce electrode impedances to 120–200 k Ω at 1 kHz.

The animals were food deprived 12 h before surgery started. In four animals, anaesthesia was induced by first placing the animal in a closed glass box filled with isoflurane vapour and then giving the animal an intraperitoneal injection of Equithesin (pentobarbital and chloral hydrate; 1.0 ml per 250 g body weight). Two animals (15272 and 15273) were anaesthetized with isoflurane (induction chamber level of 4.0% while the rats were secured in the stereotaxic apparatus, with an air flow at 1,400 ml min⁻¹; isoflurane was then gradually reduced to 1–2% during the course of the surgery). Supplementary anaesthesia was given when breathing and reflexes changed. Local anaesthetic (Xylocain) was applied on the skin before making the incision. The hyperdrive was then implanted. The tetrodes were inserted above CA3 of the right hippocampus, with the centre of the bundle at anterior–posterior 3.8 mm and medial–lateral 3.0 mm relative to bregma. Jewellers' screws and dental cement were used to secure the hyperdrive to the skull. Two screws were connected to hyperdrive ground. All tetrodes were turned after the surgery to be sure they were in the brain.

Tetrode positions. Over the course of ~3–4 weeks, the majority of the tetrodes were lowered towards CA3 in steps of 50 μ m or less while the rat rested on a towel in a large flower pot on a pedestal. Turning was slowed down when large-amplitude theta-modulated complex-spike activity appeared in CA3 at depths of approximately 3.0 mm. The tetrode depths were tweaked to get the maximal number of simultaneously recorded CA3 cells at the start of cue switching. To maintain stable recordings, the electrodes were not moved at all on the day of recording. A few tetrodes were left in CA1; data from these tetrodes were analysed separately. Two of the tetrodes were used, respectively, to record a reference signal from the corpus callosum and an EEG signal from the stratum lacunosum-moleculare.

Behavioural training procedures. The rats were trained to collect food morsels in either of two distinct 60 cm × 60 cm enclosures with 40-cm walls located in a dark curtained environment (Supplementary Fig. 2). The boxes rested on a plexiglass plate fixed 10 cm above the floor of the room. The boxes were identical except for the arrangement of a number of internal lights. Beneath the plexiglass plate of box A there was a panel of eight light-emitting white diodes (LEDs) organized into a circle (50 cm in diameter) and placed centrally under the plexiglass plate to be visible through the floor. The box was polarized by another LED at the upper edge of one of the walls. Box B was illuminated by a 60-cm-long array of green LEDs lining 40 cm of the upper edge of the wall opposite to the directional LED in A and 20 cm of one of the adjacent walls (Supplementary Fig. 13a). The LEDs were the only light source in the room.

Training occurred in four stages. During stage 1, the boxes were located next to each other, connected by a 20 × 20 cm (width × length) passageway that allowed the rat to shuttle between the boxes in order to associate each box with a different set of path integrator coordinates⁹. The rat was permitted to travel freely between the boxes for 20 min on at least three trials. Trials were separated by 20 min intervals, during which the animal rested on a towel in a pedestal outside of the curtains. At stage 2, the corridor was removed and the animal explored the boxes individually on alternating trials (3 trials in each). At stage 3, the boxes were replaced by a single box made of the same material and equipped with both sets of lights. The box was placed on alternating trials at the two original locations. When presented at the original position of box A, the set of LEDs defining A was switched on; when in the place of box B, the respective lights of B were active instead. Again, the rat was tested 3 times in each environment on alternating occasions. Finally, at Stage 4, the box was moved to a central location between the two original box locations. The animal received alternating trials with each set of lights (2 consecutive days, each day 3 pairs of trials). During all stages, at the start of each trial, the rat was taken from the flower pot on the pedestal outside of the curtains and placed, without disorientation, into the environment with the eyes gently covered by the experimenter's palm. Between trials, the rat rested for 20 min in the flower pot. During this period, the boxes were thoroughly washed with a wet tissue and dried.

On the test day, the rat started with a 20 min rest trial in the flower pot. The animal was then tested for 10 min in each box configuration (A and B, respectively). Then, during the third trial, after 40–60 s of baseline recording in one of the configurations (for example, A), the lights were switched instantaneously to the other configuration (for example, B). Additional 'teleportations' were performed every subsequent 40–60 s until 10 min had passed. A resting trial in the flower pot was recorded at the end. The electrodes were then lowered deeper into the CA3 and the experiment was repeated on the subsequent day in those cases where spikes with sufficient amplitudes from new cells could be recorded.

Running was motivated by small crumbles of cookies thrown into the box at 10–20 s intervals. Three types of cookies were used: vanilla, chocolate and unflavoured. Vanilla and unflavoured were given in configuration A, chocolate and unflavoured in configuration B. The proportion between flavoured and unflavoured cookies was approximately 1:1 during stage 1 and 1:4 during the following stages. During teleportation trials, only unflavoured crumbles were offered.

Population vector analyses. Theta waves were identified from filtered local EEG traces as described in the Supplementary Methods. The evolution of unit activity over successive theta cycles before and after cue switching was estimated by defining a population vector for each theta cycle, consisting of the number of spikes fired by each cell in the array, and comparing it to a population vector for the same cells at the corresponding location in each of the reference environments (Fig. 1a, c). Boundaries between theta cycles were determined by plotting, for all cells on the entire teleportation session, the number of spikes as a function of theta phase (Fig. 1b). The boundary between successive theta cycles was then defined by the theta phase with the lowest overall firing rate (bins of 10 deg; Fig. 1b, c). For each theta cycle, the momentary population vector was correlated with reference population vectors for the same cell sample at the same location in each of the reference environments. The reference vector (expressed in Hz rather than as a spike count, normalized by the time spent at each location) was based on activity across the entire reference trial for each of the environments on the test day. In a subset of the analyses, the momentary population vector was compared with all reference population vectors in the two boxes, that is, not only those corresponding to the animal's current location. Correlations between momentary population vectors and reference vectors were quantified by dot products as well as Pearson product–moment correlations. Dot products were calculated for all theta cycles, including those with no activity, where the dot product is zero. Pearson correlations, which cannot be defined for theta cycles with no spikes, were calculated by excluding also theta cycles with a single unit active, that is, these correlations were based on the subset of theta cycles that included at least two active cells. In a subset of the analyses, also cycles with only two active units were excluded, and in a further control, also those with only three active units were discarded (Supplementary Fig. 7).

To determine if the similarity between momentary population vectors during cue switching and population vectors in the reference environment was larger than expected by chance, we correlated population vectors from successive theta cycles in reference recordings from A and B with mean population vectors generated from activity in the same position across the entire session in the same environment or the alternative environment. When referenced to the same environment ($A \times A$ or $B \times B$), the distribution of Pearson correlations was centred at high correlation values with a long tail of low values (Supplementary Fig. 9, top row). When referenced to the alternative environment ($A \times B$ or $B \times A$), the distributions were centred at slightly negative values with a long tail towards infrequent high positive values (Supplementary Fig. 9, bottom row). We then determined 5th and 95th percentile values for each distribution and used these as criteria to identify flickers to the alternative representation. Individual theta cycles were defined as A-correlated if the correlation with reference environment A was above the 95th percentile for $B \times A$ (that is, more similar to A than 95% of the theta cycles in B) and if the correlation with reference environment B was below the 5th percentile for $B \times B$ (that is, more different from the reference in B than most theta cycles in B). Conversely, theta cycles were defined as B-correlated if the correlation with reference environment B exceeded the 95th percentile for $A \times B$ and if the correlation with reference environment A was lower than the 5th percentile for $A \times A$. In separate analyses, the momentary population vectors were compared not only with reference vectors at the same location but at all 900 bin locations in the box (Supplementary Fig. 8).

The abundance of theta cycles that correlate with both reference environments ('mixed' theta cycles) was determined by comparing the data with spike patterns obtained by 'shuffling' the activity of individual units. For each theta cycle during the 10 s preceding and succeeding the cue switching, the number of spikes produced by each unit recorded during the cue switching was drawn at random from among all theta cycles recorded in the same physical location in the same environment during the corresponding period (most of these theta cycles were correlated with the current environment, some with the alternative one). The analysis was

limited to theta cycles for which a minimum of 3 theta cycles had been recorded in that particular location in the relevant period, with 'same' location defined as the same $6\text{ cm} \times 6\text{ cm}$ spatial bin of the same recording box. The shuffled population vectors can be conceived as approximating the linear combination $\mathbf{r}_S \approx \mathbf{A} \cdot \mathbf{r}_A + \mathbf{B} \cdot \mathbf{r}_B$ where \mathbf{A} and $\mathbf{B} = 1 - \mathbf{A}$ are random binary vectors (for example, (0,1,0,0,1,0,1,1,...)) indicating whether each unit was drawn from the A or B representation of that location. If the current environment is A, a particular theta cycle in the original data can express activity close to the reference vector A in that location, or to the reference vector B (for example, during a flicker event) or to a mixture of the two, or just noise or, rather frequently, no activity at all. The abundance of mixed theta cycles was determined by the un-normalized dot product DP of the population vector in each theta cycle with the reference vectors in A and B, $DP_A = \mathbf{r} \cdot \mathbf{r}_A / C$ and $DP_B = \mathbf{r} \cdot \mathbf{r}_B / C$ (just divided by the number C of simultaneously recorded units). DP values range from 0 to about 10 (note that they are expressed in Hz, as the momentary population vector includes spike counts whereas the reference vector indicates firing rates), and very rarely beyond, but are mostly clustered close to zero because of the presence of theta cycles with no

spikes or with, for example, a single spike from a unit that emits on average less than 1 spike in that physical location in both reference vectors. We arbitrarily set $DP > 1$ as the criterion for substantial correlation, and $DP > 0.5$ for half-cycles.

The speed of instantaneous transitions from one representation (for example, A) to the other (for example, B) was estimated by cross-correlating the sequence of correlations with reference environment A and the sequence of correlations with B one time-bin later, excluding, in this analysis, units with overlapping fields in the two environments and trials with less than 20 non-overlapping units (2 out of 11; Supplementary Fig. 3). Bin widths were then varied from 44 ms to 500 ms to determine the transition time that gave the largest cross-correlation. Bins were also defined by actual theta cycles, with separations at the point of minimum spike activity in the population (Fig. 1b, c) and at other phases 10 deg apart (Supplementary Fig. 3). Cross-correlation values were corrected for overlap between A and B representations by subtracting the cross-correlation at zero time lag.

Other. Recording procedures, criteria for spike sorting, construction of rate maps, analysis of theta rhythm and histological procedures are described in Supplementary Materials and Methods.

Corridors of migrating neurons in the human brain and their decline during infancy

Nader Sanai^{1,2,3}, Thuhien Nguyen¹, Rebecca A. Ihrle¹, Zaman Mirzadeh^{1,3}, Hui-Hsin Tsai¹, Michael Wong¹, Nalin Gupta², Mitchel S. Berger², Eric Huang⁴, Jose-Manuel Garcia-Verdugo⁵, David H. Rowitch^{1,2,6} & Arturo Alvarez-Buylla^{1,2}

The subventricular zone of many adult non-human mammals generates large numbers of new neurons destined for the olfactory bulb^{1–6}. Along the walls of the lateral ventricles, immature neuronal progeny migrate in tangentially oriented chains that coalesce into a rostral migratory stream (RMS) connecting the subventricular zone to the olfactory bulb. The adult human subventricular zone, in contrast, contains a hypocellular gap layer separating the ependymal lining from a periventricular ribbon of astrocytes⁷. Some of these subventricular zone astrocytes can function as neural stem cells *in vitro*, but their function *in vivo* remains controversial. An initial report found few subventricular zone proliferating cells and rare migrating immature neurons in the RMS of adult humans⁷. In contrast, a subsequent study indicated robust proliferation and migration in the human subventricular zone and RMS^{8,9}. Here we find that the infant human subventricular zone and RMS contain an extensive corridor of migrating immature neurons before 18 months of age but, contrary to previous reports⁸, this germinal activity subsides in older children and is nearly extinct by adulthood. Surprisingly, during this limited window of neurogenesis, not all new neurons in the human subventricular zone are destined for the olfactory bulb—we describe a major migratory pathway that targets the prefrontal cortex in humans. Together, these findings reveal robust streams of tangentially migrating immature neurons in human early postnatal subventricular zone and cortex. These pathways represent potential targets of neurological injuries affecting neonates.

We collected human brain specimens from 10 neurosurgical resections and 50 autopsied brains, ranging in age from birth to 84 years (Supplementary Fig. 1 and Supplementary Table 1), using a protocol that allows subsequent analysis by fluorescent immunohistochemistry and *in situ* hybridization (see Methods). As shown (Fig. 1), staining of horizontal sections (30 μ m) through the anterior horn of the lateral ventricle demonstrates that, in the first 6 months of life, the structure of the human subventricular zone in infants differs considerably from that observed in adults. The astrocytic ribbon and gap layer are not evident (Fig. 1) and, as seen in the fetal human brain^{10,11}, cells with elongated radial glial processes line the lateral ventricular wall and express vimentin and glial fibrillary acidic protein (GFAP)¹¹. Adjacent to these radial glia, we observed a dense network of elongated unipolar and bipolar cells oriented tangentially to the ventricular lining. Many of these cells expressed the immature neuronal markers doublecortin (DCX) (Fig. 1) and β -III tubulin (TUBB3). Some putative immature neurons also expressed polysialylated neural cell adhesion molecule (PSA-NCAM), which is present in migratory cells. They also had ultrastructural features of immature migrating neurons (Supplementary Fig. 2) similar to neuroblasts described in the rodent subventricular zone¹². Progressively, between 6 to 18 months of age, the subventricular zone is depleted of this dense network of putative migratory neurons

and adopts the characteristic adult structure with an astrocyte ribbon and hypocellular gap layer. The emergence of the gap layer coincides with the decline in DCX⁺ immature neurons (25-fold during the first 6 months; $n = 16$, ages 0–17 years; Fig. 1) and proliferation, suggesting that human subventricular zone neurogenesis decreases drastically during the first 6 months of life. Only a small number of proliferating cells were present in adolescents and adults. Expression of the proliferation marker Ki67 was not associated with pyknotic nuclei, although we cannot exclude that some Ki67⁺ nuclei correspond to apoptotic cells induced by ischaemia¹³. We also identified a subpopulation of epidermal growth factor receptor (EGFR)-positive cells, a marker associated with early progenitors including neural stem cells¹⁴ and transit-amplifying cells in mice¹⁵, which similarly diminished with time (Supplementary Fig. 3). A subset of EGFR⁺ cells expressed Ki67, while others co-localized with DCX or PSA-NCAM, potentially representing transitional stages from transit-amplifying cells to immature neurons. Whole-mount *en face* preparations of the subventricular zone also confirmed massive numbers of tangentially oriented DCX⁺ chains within the gap layer at 1 week and 2 months of life. In stark contrast, cells with the morphology and marker expression of immature migratory neurons were extremely rare in adults (Supplementary Fig. 4).

Our results indicate that robust streams of tangentially migrating immature neurons initially populate the postnatal human gap layer; these pathways become depleted between 6–18 months and this region transitions into a hypocellular gap. Previous work has also demonstrated a similarly sharp decline in EGFR and PSA-NCAM immunoreactivity in the human subventricular zone during the first year of life¹⁷. Within the subventricular zone, the decline of migratory immature neurons appears first along the posterior third of the lateral ventricle and then progresses in a posterior-to-anterior trajectory towards the ventral tip (data not shown). After 18 months, Ki67⁺ proliferative activity and the number of DCX⁺ immature neurons assume trace levels seen in adults, leaving behind the gap layer characteristic of the adult human subventricular zone⁷.

We next investigated whether proliferation and the presence of putatively migrating, immature neurons in the subventricular zone were associated with an active RMS. Using autopsied material ($n = 6$) (ages 1 day; 1 week; 1, 3 and 6 months), serial sagittal and coronal reconstruction of the ventral forebrain revealed an uninterrupted column of cells that connected the ventral tip of the subventricular zone to the olfactory peduncle (Fig. 2). The descending/proximal limb of the paediatric RMS contained cells organized as chains—large collections of elongated immature neurons expressing DCX and surrounded by glial cells and processes^{1,4,18,19} (Fig. 2)—or as broad streams of individual cells. A subpopulation of these DCX⁺ cells expressed PSA-NCAM. Conversely, analysis of tissue from older children ($n = 7$; 2, 3, 7, 16, 17 years) failed to reveal chains of migrating cells or evidence of an active RMS. However, individual or pairs of elongated DCX⁺ PSA-NCAM⁺

¹Eli and Edythe Broad Institute of Regeneration Medicine and Stem Cell Research and Howard Hughes Medical Institute, University of California San Francisco, San Francisco, California 94143, USA.

²Department of Neurological Surgery, University of California San Francisco, San Francisco, California 94143, USA. ³Barrow Brain Tumor Research Center, Barrow Neurological Institute, Phoenix, Arizona 85013, USA. ⁴Department of Pathology, University of California San Francisco, San Francisco, California 94143, USA. ⁵Laboratorio de Morfología Celular, Unidad Mixta CIPF-UVeG, CIBERNED, Valencia 46012, Spain. ⁶Department of Pediatrics, University of California San Francisco, San Francisco, California 94143, USA.

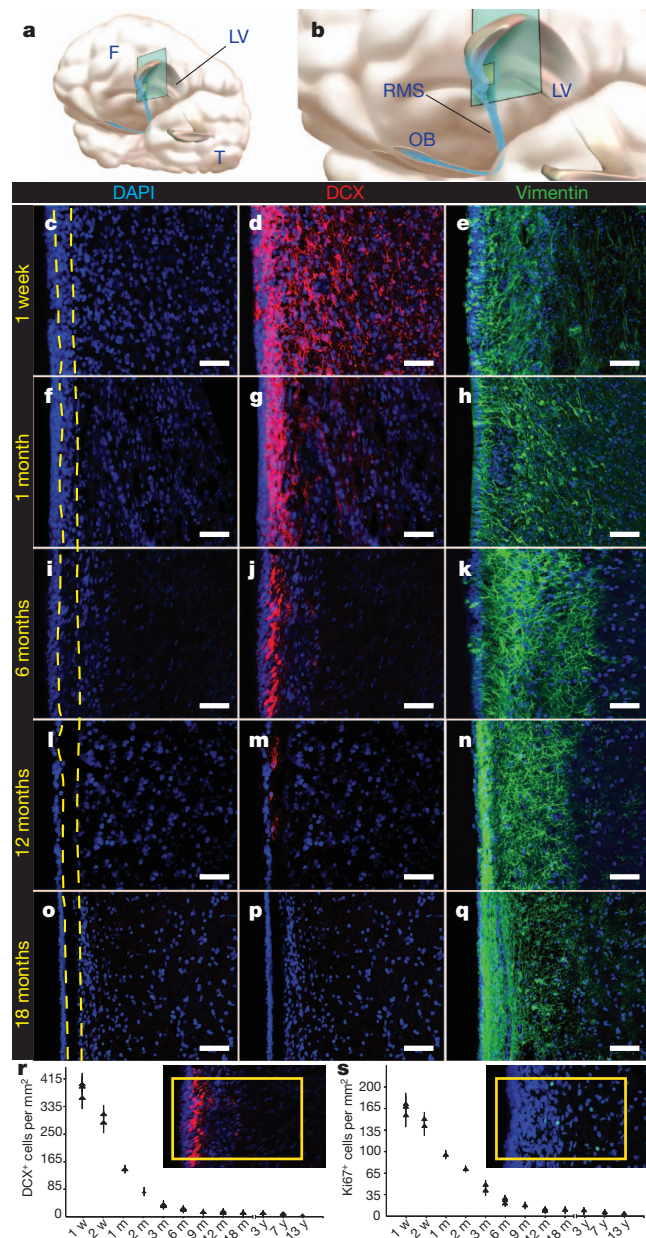


Figure 1 | Cytoarchitectural development of the human subventricular zone during the first 18 months of life. **a–q**, Illustrations localizing the horizontal 30- μ m sections of the anterior ventral subventricular zone (**a**, **b**) labelled with nuclear marker DAPI (**c**, **f**, **i**, **l**, **o**), the immature migrating neuronal marker DCX (**d**, **g**, **j**, **m**, **p**) and the immature glial marker vimentin (**e**, **h**, **k**, **n**, **q**). **F**, frontal lobe; **LV**, lateral ventricle; **OB**, olfactory bulb; **T**, temporal lobe. Large numbers of DCX⁺ immature migrating neurons populate the region that develops into the human gap layer (yellow dotted lines). **e**, **h**, **k**, **n**, **q**. In the first few months of life, radial glia-like vimentin⁺ cells are seen to populate the ventricular lining (**e**, **h**), and at 12 months onward, a dense network of vimentin⁺ processes fill the gap area (**n**, **q**). DCX and DAPI are co-labelled in the same section (first and second columns), whereas vimentin staining is from adjacent sections (third column). **r**, DCX⁺ immature neurons within the anterior ventral subventricular zone were quantified in ten $\times 20$ fields per section, ten sections per specimen, and 1–3 specimens per time point. Each data point represents the mean \pm s.d. of a single specimen at the designated age (m, month; w, week; y, year). Inset demonstrates the field of quantification (yellow box) for each of the ten horizontal sections per specimen. **s**, Proliferating Ki67⁺ subependymal cells were similarly quantified along the anterior ventral human subventricular zone. Taken together, these data demonstrate, during the first 18 months of life, a sharp decline in proliferating cells and immature migrating neurons coinciding with emergence of the human subventricular zone gap layer. Scale bars, 30 μ m.

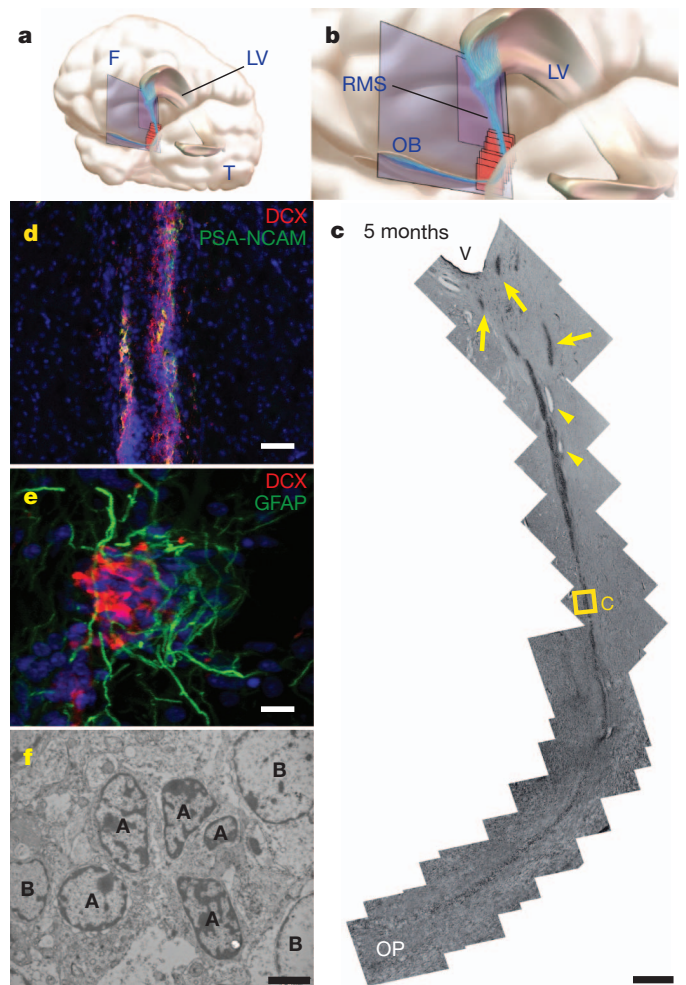


Figure 2 | The infant human RMS connects the subventricular zone to the olfactory peduncle. **a**, **b**, Illustrations indicating the anatomical site, plane and orientation of tissue sectioning. **F**, frontal lobe; **LV**, lateral ventricle; **OB**, olfactory bulb; **T**, temporal lobe. **c**, Reconstruction of serial sagittal sections of a haematoxylin-stained RMS from inferior lateral ventricle (**V**) to olfactory peduncle (**OP**) of an infant specimen (5 months of age). The subventricular source of the proximal RMS shows multiple tributaries of cells (yellow arrows) that appear to coalesce forming the descending (proximal) limb of the RMS. There is no evidence of a continuous ventricular system, but isolated ependymal islets (yellow arrowheads) are observed along the RMS. The distal limb of the RMS moves laterally, out of the sagittal plane of sectioning, as indicated by the yellow section frames. Yellow box (**C**) indicates region of proximal limb of RMS shown at high resolution in **d**. **d**, Clusters of DCX⁺ PSA-NCAM⁺ immature neurons at high magnification. **e**, Individual DCX⁺ neuronal chains, seen here in the ventral tip of the anterior horn of a 1-week specimen, are typically embedded in a matrix of GFAP⁺ glial processes and often migrate along blood vessels (see Supplementary Fig. 4d). **f**, Ultrastructural analysis of a neuronal chain cross-section identifies immature neurons (type A cells, labelled 'A') surrounded by glia (type B cells, labelled 'B'). Scale bars, 500 μ m (**c**), 50 μ m (**d**), 10 μ m (**e**) and 2 μ m (**f**).

putative migratory neurons were occasionally noted in late childhood specimens ($n = 2$; 3 and 7 years) and adult specimens ($n = 5$, ages 30, 41, 61, 74, 84 years). These observations indicate that, whereas the infant brain contains a robust RMS with massive chain migration, such activity is reduced markedly in older children and adults⁷. Although our data do not rule out that rare immature neurons may sporadically migrate within the subventricular zone and RMS at later stages^{7,9}, these findings do not support the previous finding of robust proliferation and abundant migration within the adult human subventricular zone⁸.

The distal limb of the RMS delivers subventricular zone neuronal progeny to the human olfactory peduncle, and then to the olfactory tract and olfactory bulb. Serial cross-sections of the olfactory tract ($n = 6$; 1 day; 1 week; 5, 6 and 8 months) revealed clusters of DCX⁺ PSA-NCAM⁺ immature neurons within the V-shaped central core (Fig. 3). Electron microscopy analysis of the 6-month olfactory tract ($n = 3$) confirmed ultrastructural features of migrating immature neurons within the core. Cross-sections of the proximal olfactory tract core (5 mm anterior to the olfactory peduncle) at 8 months contained 234 ± 26 total cells and 173 ± 24 DCX⁺ cells per section, respectively, while the distal olfactory tract core (5 mm posterior to the olfactory bulb) contained 216 ± 20 total cells and 115 ± 14 DCX⁺ cells. In contrast, DCX⁺ neurons were not detected in older olfactory tract specimens ($n = 3$; 18 months, 7 and 13 years) (Fig. 3). These data further support that an active RMS exists in early childhood, but is greatly reduced after 18 months of age.

Although it has been suggested that an open olfactory ventricle persists into adult life⁸, we found no evidence of a ventricular extension in the olfactory tract at any of the ages studied, consistent with previous data indicating that the olfactory ventricle fuses before birth^{7,9,20}. Along the proximal limb of the RMS, we observed ependymal islets, but no evidence of a continuous open ventricle. These displaced and discontinuous islets were lined by multiciliated cuboidal cells expressing the ciliary marker acetylated-tubulin (Supplementary Fig. 5). In all studied specimens ($n = 23$), the anterior horn of the lateral ventricle was open, but no continuous, ependymal-lined lumen extending into the RMS or olfactory tract was observed. Thus, we infer that the human olfactory tract serves as a conduit for neuronal chain migration to the olfactory bulb, but these chains of migratory cells are only evident in infants and occur in the absence of a ventricular extension.

In tracing the ependymal islets of the paediatric human RMS with serial coronal reconstructions ($n = 3$), we noted a decrease in calibre of the RMS from the proximal to distal limb (Fig. 4). On the basis of quantification of 4',6-diamidino-2-phenylindole (DAPI) nuclear staining, the proximal limb of the RMS contained 548 ± 66 total cells

per cross-section and 485 ± 58 DCX⁺ immature neurons per cross-section. In the distal limb, however, there were 228 ± 24 total cells per cross-section and 189 ± 17 DCX⁺ immature neurons per cross-section, equating to a 58% decline in total cells and a 61% decline in immature neurons. The decreasing calibre of the RMS could be due to cell death, increased migratory speed, or immature migratory neurons taking alternative paths. TdT-mediated dUTP nick end labelling (TUNEL)-positive cells were present in the proximal and distal RMS, raising the possibility that apoptosis could contribute to this decline. However, serial coronal reconstruction of the frontal lobe also unexpectedly revealed an additional migratory stream of DCX⁺ cells branching off the proximal limb of the RMS and ending in the ventromedial prefrontal cortex (VMPFC) (Fig. 4). This medial migratory stream (MMS) was observed in human specimens aged 4–6 months but not 8–18 months. Similar to the RMS, the MMS contains large clusters of DCX⁺ and PSA-NCAM⁺ cells with elongated morphologies, some adjacent to discontinuous ependymal islets (Supplementary Fig. 5). Early time points (<1 month) also revealed a more diffuse pattern of medially oriented migratory neurons emanating from the proximal limb of the primary RMS (Supplementary Fig. 6). Interestingly, cells expressing DCX, PSA-NCAM and the interneuron markers calretinin (CalR) and tyrosine hydroxylase (TH), were observed not only within this MMS but also within a restricted subregion of the VMPFC (Fig. 4). In contrast, very few DCX⁺, PSA-NCAM⁺, CalR⁺ and TH⁺ cells were evident in adjacent areas of prefrontal cortex. Although it remains possible that alterations in immunoreactivity allow progeny to escape into adjacent regions undetected, these observations suggest that the MMS diverts immature subventricular zone neurons to target the VMPFC.

A comparable MMS has not been reported in other vertebrates. We analysed serial sections of mouse brains at postnatal day (P)4, P8, P16 and P20 and did not detect a MMS (Supplementary Fig. 7). However, some individual DCX⁺ cells were observed migrating ventrally and laterally in juvenile brain studied that might target homologous brain regions. In rodents and non-human primates, migratory neurons can

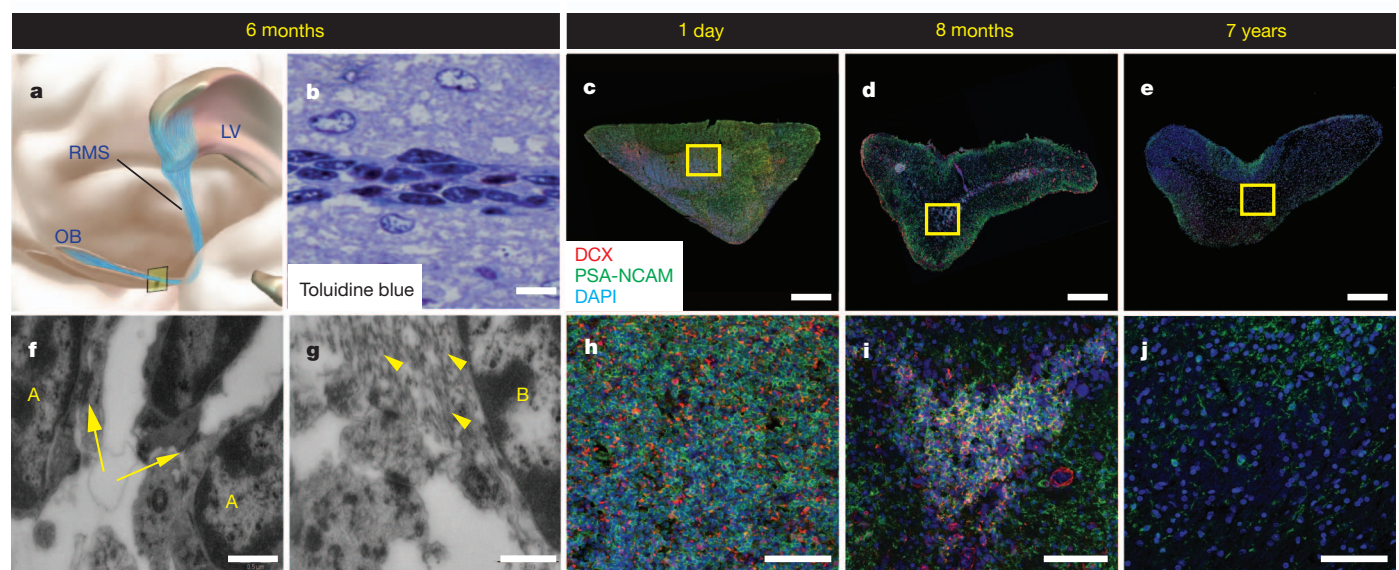


Figure 3 | Postnatal development and decline of the RMS in the human olfactory tract. **a, c–e, h–j.** Illustration (**a**) localizes the anatomical sites for cross-sectional reconstructions of the proximal human olfactory tract at 1 day (**c, h**), 8 months (**d, i**) and 7 years (**e, j**). LV, lateral ventricle; OB, olfactory bulb. High-magnification fields (**h–j**) demonstrate progressive loss of DCX⁺ PSA-NCAM⁺ immature neurons in the olfactory tract. An active RMS is evident at birth and 8 months of age, with many DCX⁺ PSA-NCAM⁺ cells. By 7 years, however, chains of migrating cells are not observed and DCX⁺ cells are rare. **b.** Toluidine blue staining of a longitudinal semi-thin (6- μ m) section from a

6-month olfactory tract demonstrates a central chain of darkened immature neurons surrounded by a matrix of lighter-coloured astrocytes. **f, g.** Electron microscopy also demonstrates clusters of type A cells (labelled 'A') enmeshed with type B cells within the olfactory tract core, including intercellular junctions (yellow arrows) between immature neurons characteristic of chain migration (**f**) and intermediate filaments (yellow arrowheads) within surrounding type B cells (labelled 'B'; **g**). Scale bars, 500 μ m (**c–e**); 75 μ m (**h–j**); 10 μ m (**b**); and 0.5 μ m (**f, g**).

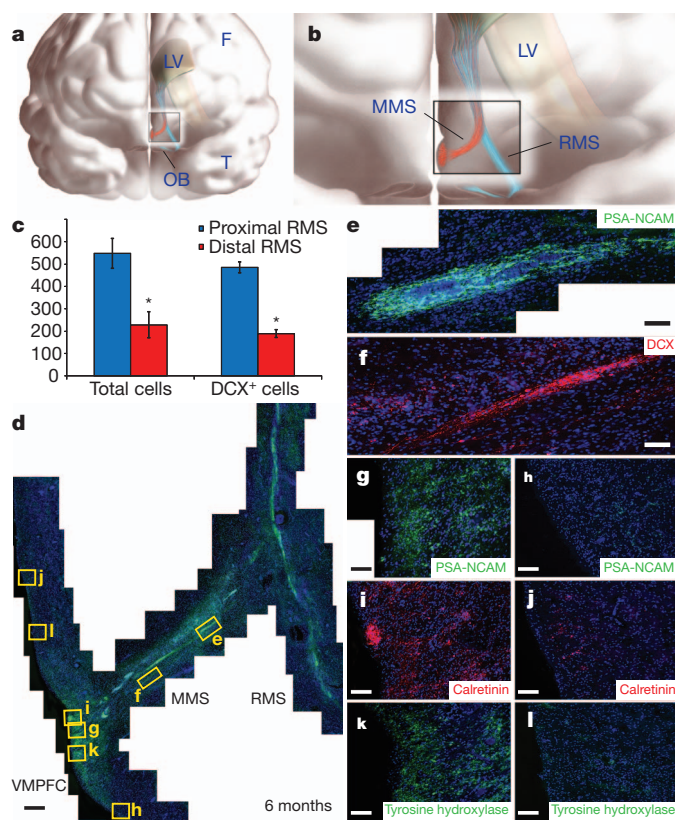


Figure 4 | A MMS of immature neurons branches from the proximal RMS in the infant human brain to supply the VMPFC. **a, b,** Illustrations demonstrate the anatomical site of the RMS (blue) and MMS (red). F, frontal lobe; LV, lateral ventricle; OB, olfactory bulb; T, temporal lobe. **c,** Using cross-sections of the proximal and distal RMS, quantifications of the total number of cells and total number of DCX⁺ cells per cross-section indicated a significant reduction from proximal to distal RMS. Asterisks indicate $P < 0.05$. Data are shown as mean \pm s.e.m. **d,** Coronal reconstruction of a 6-month specimen reveals a PSA-NCAM⁺ MMS diverging from the RMS to reach the VMPFC. **e, f,** Within this MMS, chains of PSA-NCAM⁺ cells (**e**) co-express the immature neuronal marker DCX (**f**). **g–l,** Dense clusters of PSA-NCAM⁺ (**g**), calretinin⁺ (**i**) and TH⁺ (**k**) cells are specifically observed within a subregion of the VMPFC, but not at adjacent cortical sites (**h, j, l**) superior or inferior to this subregion. Scale bars, 150 μ m (**d**), 20 μ m (**e–l**).

also escape the subventricular zone along a sagittal plane to reach the islands of Calleja^{21,22}.

Our study indicates that the region of the subventricular zone around the anterior lateral ventricles in the infant human brain is highly active, producing many tangentially migrating immature neurons. On the basis of the presence of an RMS containing chains of immature neurons, we infer that at least some of these progeny are destined for the olfactory bulb. Beyond 18 months of age, both proliferative activity and cells expressing markers of immature neurons are largely depleted, coinciding with the appearance of a hypocellular gap in the postnatal human subventricular zone. Thus, this layer of the subventricular zone initially serves as a thoroughfare for immature neurons. Surprisingly, paediatric human subventricular zone neurogenesis also seems to serve regions other than the olfactory bulb, as evidenced by medially escaping immature neurons near the RMS. These groups of cells form a unique MMS targeting a subregion of the human prefrontal cortex.

Previous work has suggested that postnatal neurogenesis may be important for learning and memory^{23,24} and that induction of plasticity may be closely linked to the timing of neuronal maturation^{25–27}. We speculate that the MMS, and other potential escape pathways from the subventricular zone, could supply interneurons to regions of the developing human brain as a mechanism of delayed postnatal

plasticity. Although the function of this recipient cortical domain, the VMPFC, is unknown in children, this region in the adult human brain is activated during specific cognitive tasks^{28,29} including spatial conceptualization and the emotional processing of visual cues. Interestingly, the VMPFC is also focally inactivated in patients with advanced Alzheimer's disease³⁰. Beyond its functional implications, this developmental study of the human subventricular zone suggests a major period of neurogenesis and neuronal migration that extends well into postnatal life, but is largely limited to early childhood. This may hold important implications for our understanding of neonatal neurological diseases, including germinal matrix haemorrhages and perinatal hypoxic–ischaemic injuries, each potentially altering subventricular zone neurogenesis and its apparent downstream cortical targets at formative stages of human development. Perhaps most importantly, the detection of a new migratory route for immature neurons within the infant human brain also highlights mechanisms through which increased regional complexity may be achieved during brain evolution. *Note added in proof:* A recent study³¹ has found the RMS at fetal stages, but very few DCX-positive neurons in this same region in the adult human brain.

METHODS SUMMARY

Human specimens. Neurosurgical excisions of normal subventricular zone occurred as part of the planned margin of resection surrounding a periventricular lesion (Supplementary Table 1). Intra-operative specimens were histologically normal with no evidence of dysplasia, and assessments were independently confirmed by an independent neuropathologist. For pathological specimens, autopsied brains were cut coronally at the mammillary bodies and immersed in 4% paraformaldehyde (PFA) for 1–2 weeks at 4 °C, and then stored in 0.1 M PBS. All specimens were collected with informed consent and in accordance with the University of California San Francisco Committee on Human Research (IRB no. H11170-19113).

Rodent specimens. Postnatal mice at the specified ages were transcardially perfused with 0.9% saline and 4% PFA, and dissected brains were post-fixed for 30 min in 4% PFA before vibratome sectioning. One-hundred-micrometre floating sections were stained and imaged using the tile scanning and three-dimensional projection modules on a Leica SP5 confocal microscope.

Immunohistochemistry. Tissue sections were incubated with primary antibodies diluted overnight at 4 °C. Sections were then incubated for 2.5 h in secondary antibodies and then incubated in streptavidin-horseradish peroxidase for 30 min. Antigen retrieval with proteinase K (10 μ g ml^{−1}), or 0.01 M citrate buffer at 95 °C was used when necessary.

Whole-mount dissection. Specimens were fixed in 4% PFA for 3 days, then rinsed in 0.1 M PBS. The anterior horn of the lateral ventricles was excised (3 mm squares) and rinsed. Tissue specimens were incubated in primary antibodies diluted in blocking solution for 2 nights. This step was repeated for secondary antibodies incubation. The ventricular face of tissue specimens was microdissected and collected at 200–300 μ m thickness and mounted.

Electron microscopy. Specimens fixed in 2% glutaraldehyde and 2% paraformaldehyde were cut into 200-nm sections on a vibratome. Sections were post-fixed in 2% osmium, rinsed, dehydrated and embedded in Araldite (Durcupan, Fluka). To identify individual cell types, ultra-thin (0.05- μ m) sections were cut with a diamond knife, stained with lead citrate and examined under a Jeol 100CX electron microscope.

Statistical measures. Cell counts were quantified and expressed as mean \pm standard deviation (s.d.) or standard error of the mean (s.e.m.). Statistical significance was determined by Student's *t*-test.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 26 March; accepted 8 August 2011.

Published online 28 September 2011.

- Kornack, D. R. & Rakic, P. The generation, migration, and differentiation of olfactory neurons in the adult primate brain. *Proc. Natl Acad. Sci. USA* **98**, 4752–4757 (2001).
- Blakemore, W. F. & Jolly, R. D. The subependymal plate and associated ependyma in the dog. An ultrastructural study. *J. Neurocytol.* **1**, 69–84 (1972).
- Pérez-Martin, M. *et al.* Ependymal explants from the lateral ventricle of the adult bovine brain: a model system for morphological and functional studies of the ependyma. *Cell Tissue Res.* **300**, 11–19 (2000).

4. Ponti, G., Aimar, P. & Bonfanti, L. Cellular composition and cytoarchitecture of the rabbit subventricular zone and its extensions in the forebrain. *J. Comp. Neurol.* **498**, 491–507 (2006).
5. Pencea, V., Bingaman, K. D., Freedman, L. J. & Luskin, M. B. Neurogenesis in the subventricular zone and rostral migratory stream of the neonatal and adult primate forebrain. *Exp. Neurol.* **172**, 1–16 (2001).
6. Young, K. M., Fogarty, M., Kessaris, N. & Richardson, W. D. Subventricular zone stem cells are heterogeneous with respect to their embryonic origins and neurogenic fates in the adult olfactory bulb. *J. Neurosci.* **27**, 8286–8296 (2007).
7. Sanai, N. *et al.* Unique astrocyte ribbon in adult human brain contains neural stem cells but lacks chain migration. *Nature* **427**, 740–744 (2004).
8. Curtis, M. A. *et al.* Human neuroblasts migrate to the olfactory bulb via a lateral ventricular extension. *Science* **315**, 1243–1249 (2007).
9. Sanai, N., Berger, M. S., Garcia-Verdugo, J. M. & Alvarez-Buylla, A. Comment on “Human neuroblasts migrate to the olfactory bulb via a lateral ventricular extension”. *Science* **318**, 393 (2007).
10. Guerrero-Cázares, H. *et al.* Cytoarchitecture of the lateral ganglionic eminence and rostral extension of the lateral ventricle in the human fetal brain. *J. Comp. Neurol.* **519**, 1165–1180 (2011).
11. Zecevic, N. Specific characteristic of radial glia in the human fetal telencephalon. *Glia* **48**, 27–35 (2004).
12. Wichterle, H., Garcia-Verdugo, J. M. & Alvarez-Buylla, A. Direct evidence for homotypic, glia-independent neuronal migration. *Neuron* **18**, 779–791 (1997).
13. Kuan, C. Y. *et al.* Hypoxia-ischemia induces DNA synthesis without cell proliferation in dying neurons in adult rodent brain. *J. Neurosci.* **24**, 10763–10772 (2004).
14. Pastrana, E., Cheng, L. C. & Doetsch, F. Simultaneous prospective purification of adult subventricular zone neural stem cells and their progeny. *Proc. Natl Acad. Sci. USA* **106**, 6387–6392 (2009).
15. Doetsch, F., Petreanu, L., Caille, I., Garcia-Verdugo, J. M. & Alvarez-Buylla, A. EGF converts transit-amplifying neurogenic precursors in the adult brain into multipotent stem cells. *Neuron* **36**, 1021–1034 (2002).
16. Mirzadeh, Z., Merkle, F. T., Soriano-Navarro, M., Garcia-Verdugo, J. M. & Alvarez-Buylla, A. Neural stem cells confer unique pinwheel architecture to the ventricular surface in neurogenic regions of the adult brain. *Cell Stem Cell* **3**, 265–278 (2008).
17. Weickert, C. S. *et al.* Localization of epidermal growth factor receptors and putative neuroblasts in human subependymal zone. *J. Comp. Neurol.* **423**, 359–372 (2000).
18. Lois, C. & Alvarez-Buylla, A. Long-distance neuronal migration in the adult mammalian brain. *Science* **264**, 1145–1148 (1994).
19. Rodríguez-Pérez, L. M., Perez-Martin, M., Jimenez, A. J. & Fernandez-Llebreg, P. Immunocytochemical characterisation of the wall of the bovine lateral ventricle. *Cell Tissue Res.* **314**, 325–335 (2003).
20. Humphrey, T. J. The development of the olfactory and accessory olfactory formation in human embryos and fetuses. *J. Comp. Neurol.* **73**, 431–468 (1940).
21. Bedard, A., Levesque, M., Bernier, P. J. & Parent, A. The rostral migratory stream in adult squirrel monkeys: contribution of new neurons to the olfactory tubercle and involvement of the antiapoptotic protein Bcl-2. *Eur. J. Neurosci.* **16**, 1917–1924 (2002).
22. Meyer, G., Gonzalez-Hernandez, T., Carrillo-Padilla, F. & Ferres-Torres, R. Aggregations of granule cells in the basal forebrain (islands of Calleja): Golgi and cytoarchitectonic study in different mammals, including man. *J. Comp. Neurol.* **284**, 405–428 (1989).
23. Zhao, C., Deng, W. & Gage, F. H. Mechanisms and functional implications of adult neurogenesis. *Cell* **132**, 645–660 (2008).
24. Nottebohm, F. The road we travelled: discovery, choreography, and significance of brain replaceable neurons. *Ann. NY Acad. Sci.* **1016**, 628–658 (2004).
25. Boveiti, S., Veyrac, A., Peretto, P., Fasolo, A. & De Marchis, S. Olfactory enrichment influences adult neurogenesis modulating GAD67 and plasticity-related molecules expression in newborn cells of the olfactory bulb. *PLoS ONE* **4**, e6359 (2009).
26. Nissant, A., Bardy, C., Katagiri, H., Murray, K. & Lledo, P. M. Adult neurogenesis promotes synaptic plasticity in the olfactory bulb. *Nature Neurosci.* **12**, 728–730 (2009).
27. Southwell, D. G., Froemke, R. C., Alvarez-Buylla, A., Stryker, M. P. & Gandhi, S. P. Cortical plasticity induced by inhibitory neuron transplantation. *Science* **327**, 1145–1148 (2010).
28. Longe, O., Senior, C. & Rippon, G. The lateral and ventromedial prefrontal cortex work as a dynamic integrated system: evidence from fMRI connectivity analysis. *J. Cogn. Neurosci.* **21**, 141–154 (2009).
29. Szatkowska, I., Szymanska, O. & Grabowska, A. The role of the human ventromedial prefrontal cortex in memory for contextual information. *Neurosci. Lett.* **364**, 71–75 (2004).
30. Herholz, K. *et al.* Discrimination between Alzheimer dementia and controls by automated analysis of multicenter FDG PET. *Neuroimage* **17**, 302–316 (2002).
31. Wang, C. *et al.* Identification and characterization of neuroblasts in the subventricular zone and rostral migratory stream of the adult human brain. *Cell Res.* <http://dx.doi.org/10.1038/cr.2011.83> (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors are grateful to J. Agudelo and R. Romero for expert technical assistance and to K. X. Probst for illustrations. N.S. was supported by an NIH F32 NRSA postdoctoral fellowship (NS 058180). R.I. was supported by fellowships from the Damon Runyon Cancer Research Foundation (DRG1935-07) and American Association for Cancer Research/National Brain Tumor Society. R.A.I. thanks the American Association for Cancer Research for support. A.A.-B. is the Heather and Melanie Muss Endowed Chair of Neurological Surgery at UCSF. This work was supported by grants from the NIH (to A.A.-B., E.H. and D.H.R.), the Pediatric Brain Tumor Foundation of the United States and by the John G. Bowes Research Fund. D.H.R. is a Howard Hughes Medical Institute Investigator.

Author Contributions N.S. designed the study, acquired and interpreted experimental data, and prepared the manuscript. T.N. assisted with experiments, data collection and manuscript preparation. R.A.I. designed and conducted the EGFR experiments and assisted with manuscript preparation. Z.M. designed and conducted the whole-mount experiments. H.-H.T. and M.W. conducted the *in situ* hybridization experiments. N.G., M.S.B. and E.H. assisted with specimen collection and neuropathological review. J.-M.G.-V. acquired and interpreted all ultrastructural analyses and assisted with study design. D.H.R. and A.A.-B. designed the study, interpreted the data and prepared the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.A.-B. (abuylla@stemcell.ucsf.edu) or D.H.R. (rowitchd@peds.ucsf.edu).

METHODS

Human specimens. Neurosurgical excisions of normal subventricular zone occurred as part of the planned margin of resection surrounding a periventricular lesion (Supplementary Table 1). We recorded the anatomical origin of each intra-operative specimen with intra-operative neuronavigation. Intra-operative specimens were histologically normal with no evidence of dysplasia, and assessments were independently confirmed by an independent neuropathologist.

For pathological specimens, autopsied brains were cut coronally at the level of the mammillary bodies and immersed in 4% paraformaldehyde (PFA) for 1–2 weeks, and then stored in 0.1 M PBS. The brains were then cut into 1-cm plates along the coronal, axial or sagittal plane. The anterior horn of the lateral ventricle and the ventral medial aspect, which includes the gyrus rectus and olfactory tract, of the frontal cortex were excised. Autopsy specimens were obtained within 12 h of death from all individuals. All causes of death were non-neurological in origin and all patients had no evidence of intracranial disease. All specimens were collected with informed consent and in accordance with the University of California San Francisco (UCSF) Committee on Human Research (IRB no. H11170-19113).

Rodent specimens. Postnatal male mice at the specified ages were transcardially perfused with 0.9% saline and 4% PFA, and dissected brains were post-fixed for 30 min in 4% PFA before vibratome sectioning. One-hundred-micrometre floating sections were stained with rabbit anti-doublecortin (Cell Signaling), mouse anti-PSA-NCAM (Millipore) and DAPI (Sigma), mounted on Superfrost Plus slides, and imaged using the tile scanning and three-dimensional projection modules on a Leica SP5 confocal microscope. Composite images were assembled from individual $\times 10$ fields using Adobe Photoshop. All experiments were approved by the UCSF Institutional Animal Care and Use Committee (approval no. AN077716).

Immunohistochemistry. All fixed specimens were rinsed in 0.1 M PBS and then cut on a vibratome (50 μm), or cryoprotected in 30% sucrose and then cut on a cryostat (30 μm). Tissue sections were incubated with primary antibodies diluted in TNB blocking solution (0.1 M Tris-HCl, pH 7.5, 0.15 M NaCl, 0.5% blocking reagent from PerkinElmer) overnight at 4 °C. Sections were then incubated for 2.5 h in secondary antibodies diluted in TNB. Sections were then incubated in streptavidin-horseradish peroxidase for 30 min in TNB, which catalyses subsequent fluorescent conversion of tyramide substrates (all from PerkinElmer).

Antigen retrieval with proteinase K (10 $\mu\text{g ml}^{-1}$) or 0.01 M citrate buffer at 95 °C was used when necessary.

The following antibodies were used in this study: GFAP (Chemicon), vimentin (Sigma), doublecortin (Chemicon), TuJ1 (Covance), PSA-NCAM (Genbiosys), Ki67 (DAKO), calretinin (Chemicon), calbindin (Chemicon), tyrosine hydroxylase (Chemicon) and EGFR (Upstate Biotechnology).

Whole-mount dissection. Specimens were fixed in 4% PFA for 3 days, then rinsed in 0.1 M PBS. The anterior horn of the lateral ventricles was excised (3 mm squares) and rinsed in 0.1% PBS-Triton X-100. Tissue specimens were incubated in primary antibodies diluted in blocking solution (10% normal goat serum, 2% Triton X-100, 0.1 M PBS) for 2 nights. This step is repeated for secondary antibodies incubation. DAPI was used (1:500) for counterstaining and Aqua Polymount (Polysciences) for mounting. The ventricular face of tissue specimens is microdissected and collected at 200–300 μm thickness and mounted.

In situ hybridization. Probes specific to human GFAP, doublecortin and EGFR were generated by amplifying fragments from commercially available cDNA clones (Open Biosystems) or reverse-transcribed cDNA from human fetal brain total RNA (Clontech). Primer sequences for amplifying GFAP (Allen Human Brain Atlas primer ID398603) and doublecortin (primer ID399287) were obtained from Allen Institute for Brain Science (<http://www.brain-map.org/>). Forward primer 5'-AGCTCTTCGGGGAGCAGCGA-3' and reverse primer 5'-TGACAGTGGCTTCGTCTCGG-3' were used for EGFR. Amplified fragments were subcloned into pCRII (Invitrogen) and RNA probes were made subsequently using DIG RNA labelling (Roche). Sections were cut at 14 μm , mounted on Superfrost Plus slides. The sections were pre-hybridized for 2 h and hybridized with probes at 1:100 to 1:200 dilutions at 55 °C overnight. After hybridization, slides were washed and incubated with anti-DIG antibody, and developed by BM purple substrate for 24 to 48 h. Slides hybridization with GFAP and doublecortin RNA probes were subsequently immunostained with GFAP and doublecortin antibodies as previously described.

Electron microscopy. Specimens fixed in 2% glutaraldehyde and 2% paraformaldehyde were cut into 200-nm sections on a vibratome. Sections were post-fixed in 2% osmium, rinsed, dehydrated and embedded in Araldite (Durcupan, Fluka). To study subventricular zone architecture, we cut serial 1-mm semi-thin sections and stained them with 1% toluidine blue. To identify individual cell types, ultra-thin (0.05-mm) sections were cut with a diamond knife, stained with lead citrate and examined under a Jeol 100CX electron microscope.

CORRIGENDUM

doi:10.1038/nature10518

A plastidial sodium-dependent pyruvate transporter

Tsuyoshi Furumoto, Teppei Yamaguchi,
Yumiko Ohshima-Ichie, Masayoshi Nakamura,
Yoshiko Tsuchida-Iwata, Masaki Shimamura, Junichi Ohnishi,
Shingo Hata, Udo Gowik, Peter Westhoff, Andrea Bräutigam,
Andreas P. M. Weber & Katsura Izui

Nature **476**, 472–475 (2011)

In this Letter, the DNA Data Bank of Japan accession code for *Flaveria
trinervia* BASS4 should be AB522103. This has been corrected online.

Structural basis of RNA recognition and activation by innate immune receptor RIG-I

Fuguo Jiang^{1*}, Anand Ramanathan^{2*}, Matthew T. Miller¹, Guo-Qing Tang², Michael Gale Jr³, Smita S. Patel² & Joseph Marcotrigiano¹

Retinoic-acid-inducible gene-I (RIG-I; also known as DDX58) is a cytoplasmic pathogen recognition receptor that recognizes pathogen-associated molecular pattern (PAMP) motifs to differentiate between viral and cellular RNAs. RIG-I is activated by blunt-ended double-stranded (ds)RNA with or without a 5'-triphosphate (ppp), by single-stranded RNA marked by a 5'-ppp¹ and by polyuridine sequences^{2,3}. Upon binding to such PAMP motifs, RIG-I initiates a signalling cascade that induces innate immune defences and inflammatory cytokines to establish an antiviral state. The RIG-I pathway is highly regulated and aberrant signalling leads to apoptosis, altered cell differentiation, inflammation, autoimmune diseases and cancer^{4,5}. The helicase and repressor domains (RD) of RIG-I recognize dsRNA and 5'-ppp RNA to activate the two amino-terminal caspase recruitment domains (CARDs) for signalling. Here, to understand the synergy between the helicase and the RD for RNA binding, and the contribution of ATP hydrolysis to RIG-I activation, we determined the structure of human RIG-I helicase-RD in complex with dsRNA and an ATP analogue. The helicase-RD organizes into a ring around dsRNA, capping one end, while contacting both strands using previously uncharacterized motifs to recognize dsRNA. Small-angle X-ray scattering, limited proteolysis and differential scanning fluorimetry indicate that RIG-I is in an extended and flexible conformation that compacts upon binding RNA. These results provide a detailed view of the role of helicase in dsRNA recognition, the synergy between the RD and the helicase for RNA binding and the organization of full-length RIG-I bound to dsRNA, and provide evidence of a conformational change upon RNA binding. The RIG-I helicase-RD structure is consistent with dsRNA translocation without unwinding and cooperative binding to RNA. The structure yields unprecedented insight into innate immunity and has a broader impact on other areas of biology, including RNA interference and DNA repair, which utilize homologous helicase domains within DICER and FANCM.

To investigate the contributions of the individual domains of RIG-I to RNA binding, we used fluorescence anisotropy to determine the equilibrium dissociation constants (K_d) of the protein–RNA complexes. The tightest RNA affinity was observed with the helicase-RD, whereas the full-length RIG-I, helicase domain and RD bind dsRNA with a 24-fold, 8,600-fold and 50-fold weaker affinity, respectively, and all proteins bind RNA with a 1:1 stoichiometry (Table 1 and Supplementary Fig. 1). Consistent with the affinities, full-length RIG-I and helicase-RD demonstrated robust dsRNA-stimulated ATPase activity, whereas the helicase domain showed a weak ATPase activity (Table 1). Interestingly, the presence of a 5'-ppp in dsRNA does not alter the stoichiometry of RNA binding, although the 5'-ppp dsRNA binds more tightly than the 5'-OH dsRNA to full-length RIG-I.

Crystals of RIG-I helicase-RD in complex with ADP•BeF₃ and 14 base-pair palindromic dsRNA diffracted to 2.9 Å resolution (Supplementary Table 1). The dsRNA maintains an A-form helical conformation and interacts with all four domains of helicase-RD, which are arranged into a ring around the dsRNA (Fig. 1a–c). The RecA-like helicase domain (domain 1) progresses to an α -helical domain (domain 3) and into the second RecA-like helicase domain (domain 2) that ends with the RD. The notable features of the RIG-I structure are the linkers that connect the domains. Domains 1 and 3 are connected by a β -strand that is part of the parallel β -sheet in domain 2. The RD is connected to domain 2 via a prominent V-shaped linker, consisting of two α -helices that interact extensively with domains 1 and 2 with a buried surface over 1,500 Å². The V-shaped linker extends into a proline-rich (⁷⁹⁶KPKPVPD) loop that makes the final connection to the RD. Interestingly, T770 at the vertex of the V-shaped linker is one of the phosphorylation sites that regulates RIG-I signalling⁶.

Molecular surface analysis shows that one end of the dsRNA is capped by the entire helicase-RD molecule whereas the opposite end of the RNA is exposed (Fig. 1d–f). At the capped end, the 5' terminus of the dsRNA abuts the RD domain but the 3' terminus is somewhat

Table 1 | Summary of ATPase rates and affinity measurements of dsRNA binding

	ATPase		Affinity measurements*			
	RNA	ATPase activity <i>M/(M × s)</i>	RNA	Nucleotide analogue	Dissociation constant (K_d , nM)	Final anisotropy
RD			14-bp dsRNA-F	No nucleotide	2.6 ± 1.20	0.109
Helicase	14-bp dsRNA	0.89 ± 0.04	14-bp dsRNA-F	No nucleotide	436 ± 121	0.138
	No RNA	Undetectable		ADP•BeF ₃	1300 ± 300†	0.138
Helicase-RD	14-bp dsRNA	21.8 ± 0.70	14-bp dsRNA-F	No nucleotide	0.05 ± 0.02	0.213
	14-bp palindromic dsRNA	24.4 ± 0.21	5'-ppp 14-bp dsRNA-F	ADP•BeF ₃	0.02 ± 0.07	0.280
	No RNA	Undetectable		No nucleotide	0.03 ± 0.03	0.190
	14-bp dsRNA	28.2 ± 1.42	14-bp dsRNA-F	No nucleotide	1.2 ± 0.6†	0.187
Full-length RIG-I	14-bp dsRNA	28.2 ± 1.42	5'-ppp 14-bp dsRNA	No nucleotide	0.44 ± 0.27	0.183
	No RNA	Undetectable				

Mean ATPase rate constant from two independent measurements and range is shown. Affinity measurements and range from two or more independent experiments are shown. Initial anisotropy of the free fluorescein-labelled dsRNA was 0.10 for all experiments.

* Data was fitted to a 1:1 (protein:RNA) binding model

† Value from single measurement and fitting error is shown.

¹Center for Advanced Biotechnology and Medicine, Department of Chemistry and Chemical Biology, Rutgers University, 679 Hoes Lane West, Piscataway, New Jersey 08854, USA. ²Department of Biochemistry, UMDNJ-RWJ Medical School, 675 Hoes Lane West, Piscataway, New Jersey 08854, USA. ³Department of Immunology, University of Washington School of Medicine, 1959 NE Pacific Street, Seattle, Washington 98195, USA.

*These authors contributed equally to this work.

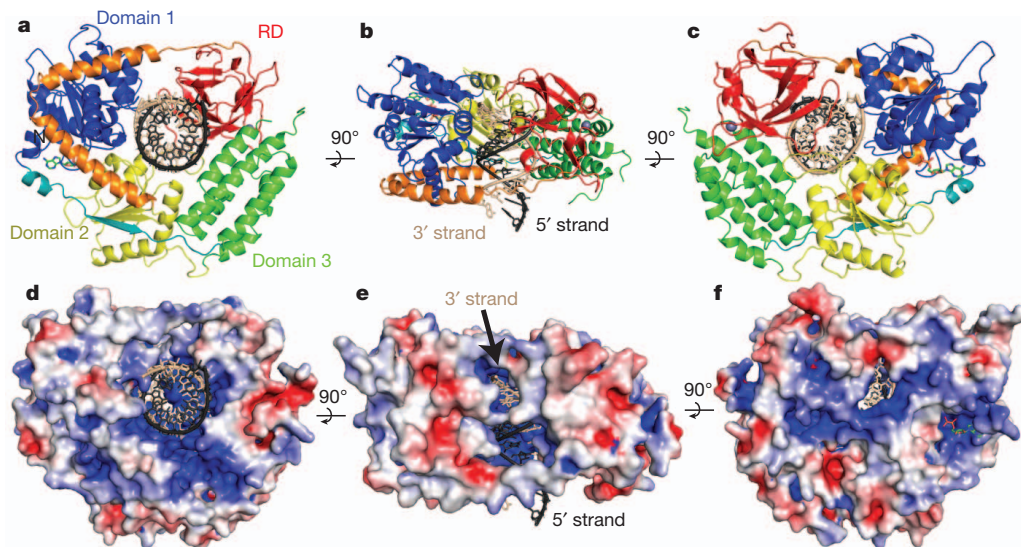


Figure 1 | Structural overview of RIG-I helicase-RD. **a–c**, Schematic representation of the RIG-I helicase-RD, highlighting the RecA-like domain 1 (blue), the α -helical domain 3 (green), the RecA-like helicase domain 2 (yellow) and the RD (red). The linker connecting domain 1 with domain 3 is coloured teal, and the V-shaped linker between domain 2 and the RD is coloured orange. The ADP•BeF₃ and dsRNA are shown in stick representation with the 5' and 3'

exposed through a highly basic channel (Fig. 1b, e). The 3'-terminus RNA strand will be referred to as the 3' strand, and the opposite strand will be referred to as the 5' strand. This basic channel may allow RIG-I

strands of the RNA coloured black and beige, respectively. A grey sphere denotes the position of the zinc ion in the RD. **d–f**, The surface of the RIG-I helicase-RD coloured for electrostatic potential at $\pm 5 \text{ kT e}^{-1}$: blue (basic), white (neutral) and red (acidic). The views in panels **a**, **b** and **c** are identical to those in **d**, **e** and **f**, respectively.

to recognize dsRNA with 3' nucleotide overhangs or 3' monophosphates that are products of RNase L digestion⁷. Additionally there are visible channels along the long axis of the dsRNA, which may allow the

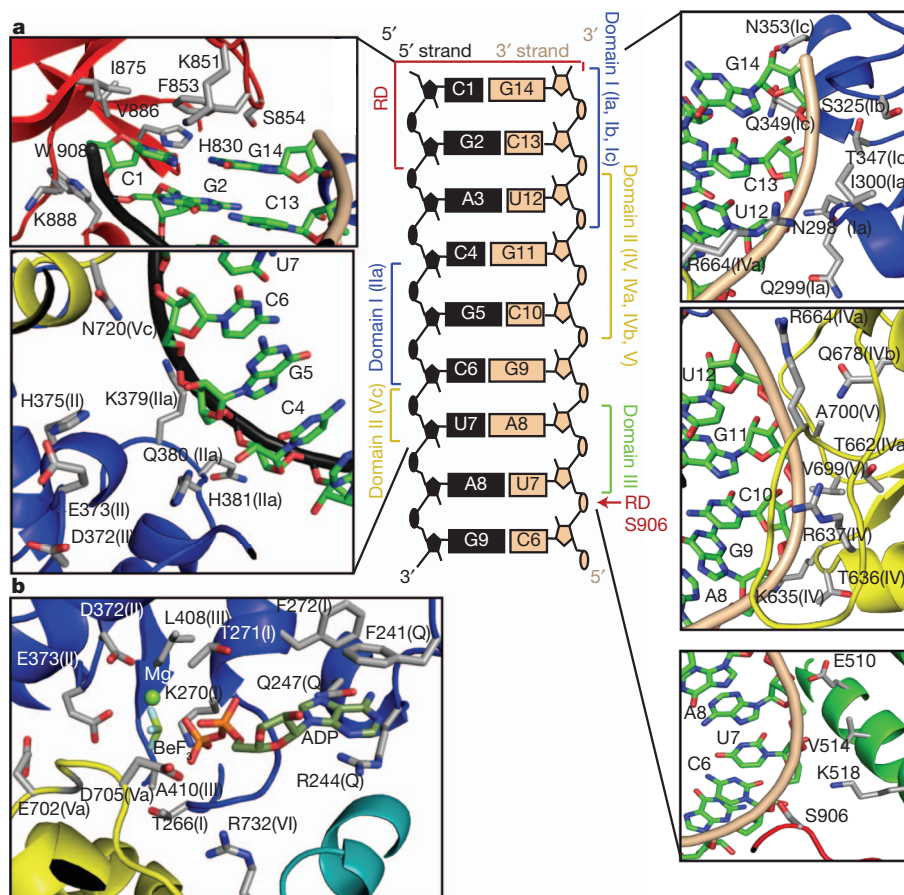


Figure 2 | Interactions of the RIG-I helicase-RD with dsRNA and ADP•BeF₃. **a**, A schematic representation showing the interactions between RIG-I domains and helicase motifs (given in parentheses) with dsRNA located

in the centre. Detailed contacts are shown in the surrounding panels. **b**, Stick representation detailing the interactions of RIG-I helicase motifs with ADP•BeF₃ and Mg²⁺ is shown in the lower left panel.

domains to flex and accommodate bulges or noncanonical base pairs (Fig. 1e, f).

All four domains of the helicase-RD participate in dsRNA binding, burying a total of 1,500 Å² of surface area to encircle about eight base pairs (Fig. 2a). Most contacts are with the sugar phosphate backbone of both strands but there are a few base-specific interactions. The F853 in the RD stacks over the terminal C1–G14 base pair, and H830 and S854 form hydrogen bonds with the ribose 2'-OH of C1 (5' strand) and G14 (3' strand), respectively. Phosphorylation of S854 and S855, shown to negatively regulate RIG-I signalling⁶, would be predicted to adversely affect RNA binding. The RD–RNA contacts in the helicase-RD structure are identical to those identified in the structures of the isolated RD bound to dsRNA^{8–10}. Residues H847, K858 and K861 of the RD, reported to interact with the 5'-ppp, do not make any new interactions in the helicase-RD structure. The only RD contact with the 3' strand, other than at the blunt end, is between S906 and the U7 phosphate backbone.

The core helicase (domains 1 and 2) contains characteristic motifs critical for RNA binding and ATP hydrolysis¹¹ (Supplementary Fig. 3). The two helicase domains together contact five nucleotides from the 3' end of the 3' strand (G14–C10) and four nucleotides in the middle of the 5' strand (C4–U7) (Fig. 2a). Motif Ib (S325 and G326) and motif Ic (T347, Q349 and N353) of domain 1 contact the ribose-phosphate backbone of the terminal base of the 3' strand (G14), while motif Ia (N298, Q299 and I300) interacts with C13 and U12. Motifs IV, IVa, IVb and V in domain 2 interact with the 3' strand from C13 to C10. Residues in motif IV (K635, T636 and R637) interact with the phosphate backbone of G11–C10. Similarly, motif IVa (T662, G663 and R664) and motif IVb (Q678) contact the backbone of C13–G11, and motif V (V699) interacts with the RNA backbone at C10.

All RIG-I-like helicases contain a large insertion (domain 3) between the core helicase domains¹². The structure of the RIG-I-like helicase Hef showed a similar α -helical domain 3, except it is rotated in RIG-I owing to dsRNA interactions (Supplementary Fig. 4). An α -helix within domain 3 (residues 506–522) of RIG-I runs almost perpendicular to the minor groove of the dsRNA and interacts with the 3' strand without contacting the 5' strand (Fig. 2a). The interactions of domain 3 with the dsRNA via residues E510, V514 and K518 extend the helicase contacts to U7 and A8 of the 3' strand.

Although most interactions of the RIG-I are with the 3' strand, all three helicase domains are in close proximity to the 5' strand and two new motifs contact the 5' strand. Motif IIa (³⁷⁹KQHPY) immediately follows motif II (³⁷²DECH) and interacts with C4–C6 of the 5' strand. Similarly, N720 in motif Vc in domain 2 contacts the 5'-strand backbone at C6 and U7. These new motifs may represent a general feature of helicases binding to dsRNA or dsDNA, as a region similar to motif IIa was identified in the Swi/Snf2 family helicases and a Rad54 homologue dsDNA complex structure¹³.

RIG-I helicase-RD contains one ADP•BeF₃ molecule bound at the interface of domains 1 and 2 via the conserved helicase motifs Q, I, II, III, Va and VI, which are generally involved in ATP binding/hydrolysis (Fig. 2b)¹². The Q motif makes adenine-specific contacts, whereas the other motifs are involved in binding the triphosphate moiety and the Mg²⁺. Motif I (²⁶⁷CGGKT) contacts the phosphates and the BeF₃, which mimics the γ -phosphate of ATP as observed in other SF2 helicases¹². The D372 and E373 of motif II (³⁷²DECH) coordinate the Mg²⁺ to stabilize the ATP analogue. The helicase motifs Va and VI, which contact the ribose and phosphates of the ATP in other helicases, are close to ADP•BeF₃.

RIG-I has been shown to have translocation activity on dsRNA¹⁴ and has been reported to unwind short dsRNA¹⁵, although we (data not shown) and others¹⁴ have failed to detect unwinding activity. Superposition of RIG-I with hepatitis C virus (HCV) NS3 helicase (NS3h) bound to single-stranded (ss)DNA¹⁶ (Fig. 3a) makes predictions about translocation and lack of helicase activity of RIG-I. First, ssDNA in NS3h overlays with the 3' strand of the dsRNA bound to

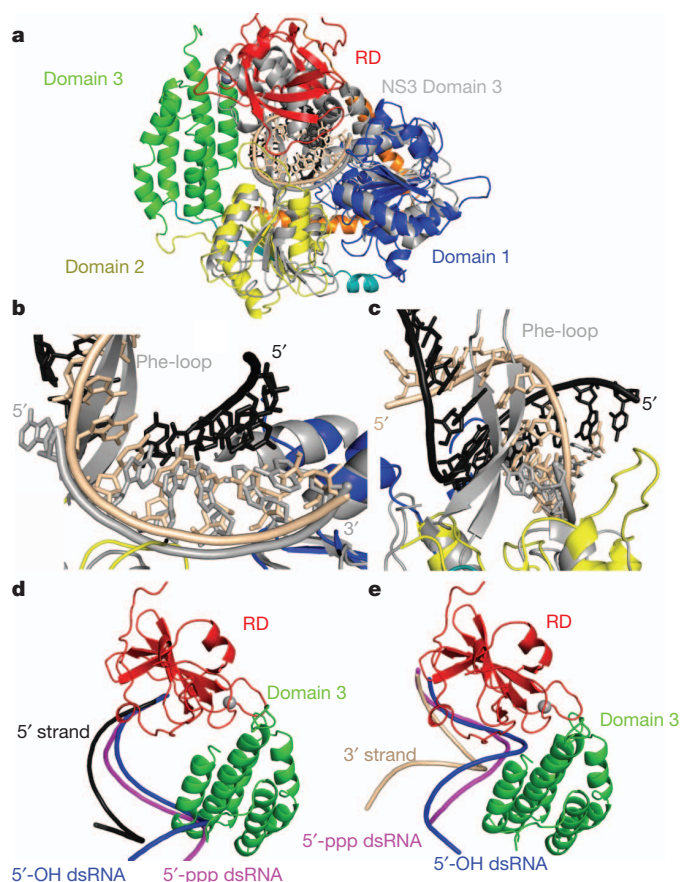


Figure 3 | Comparison of RIG-I helicase-RD with HCV NS3h and RD bound to 5'-OH and 5'-ppp dsRNA. **a**, Ribbon diagram showing the superposition of the RIG-I helicase-RD–dsRNA–ADP•BeF₃ structure and NS3h bound to ssDNA (PDB accession 3KQH) (grey). The helicase core domains 1 and 2 from RIG-I helicase-RD superimpose well, whereas domain 3 of NS3h is positioned over the RD. **b**, Superposition of the RIG-I helicase-RD with NS3h demonstrates that the ssDNA bound to NS3h overlays with the 3' strand (beige) of the dsRNA bound to the helicase-RD. **c**, The location of the Phe-loop of NS3h relative to the dsRNA of the RIG-I helicase-RD–dsRNA–ADP•BeF₃ structure. **d, e**, Superposition of the 5'-OH (blue; PDB accession 3OG8) and 5'-ppp dsRNA (magenta; PDB accession 3LRR) based on the location of the RD. For clarity the 5' strands (**d**) and 3' strands (**e**) are shown separately.

RIG-I, suggesting that the RIG-I helicase makes principal motor contacts with the 3' strand (Fig. 3b) when translocating along dsRNA. Second, a conserved Phe-loop in NS3h bisects the dsRNA, consistent with its implicated role in unwinding¹⁷ (Fig. 3b, c). The absence of such a motif in RIG-I could explain the lack of RNA-unwinding activity. Interestingly, the RIG-I RD positions over domain 3 of NS3h, indicating that this domain of NS3h may interact with the 5' strand of dsRNA (Fig. 3a).

The RD superimposes well with a root mean squared deviation of ~0.6 Å with the structures of isolated RD bound to dsRNA with and without 5'-ppp^{8–10}. However, the trajectory of the RNA helix in the helicase-RD is different. Superposition of the three RD structures demonstrates that the dsRNA bound to the isolated RD clashes with domain 3 (Fig. 3d, e). Thus, the dsRNA in helicase-RD is rotated and the previously reported contacts (R811, K849, K851 and H871) between RD and RNA are no longer observed in the helicase-RD structure.

It has been proposed that RIG-I exists in an auto-inhibited state¹⁸ with undetectable ATPase activity (Table 1) in the absence of RNA, and activates upon binding viral RNAs. Small-angle X-ray scattering (SAXS), limited proteolysis and differential scanning fluorimetry (DSF) were performed to gain insights into how RNA binding activates

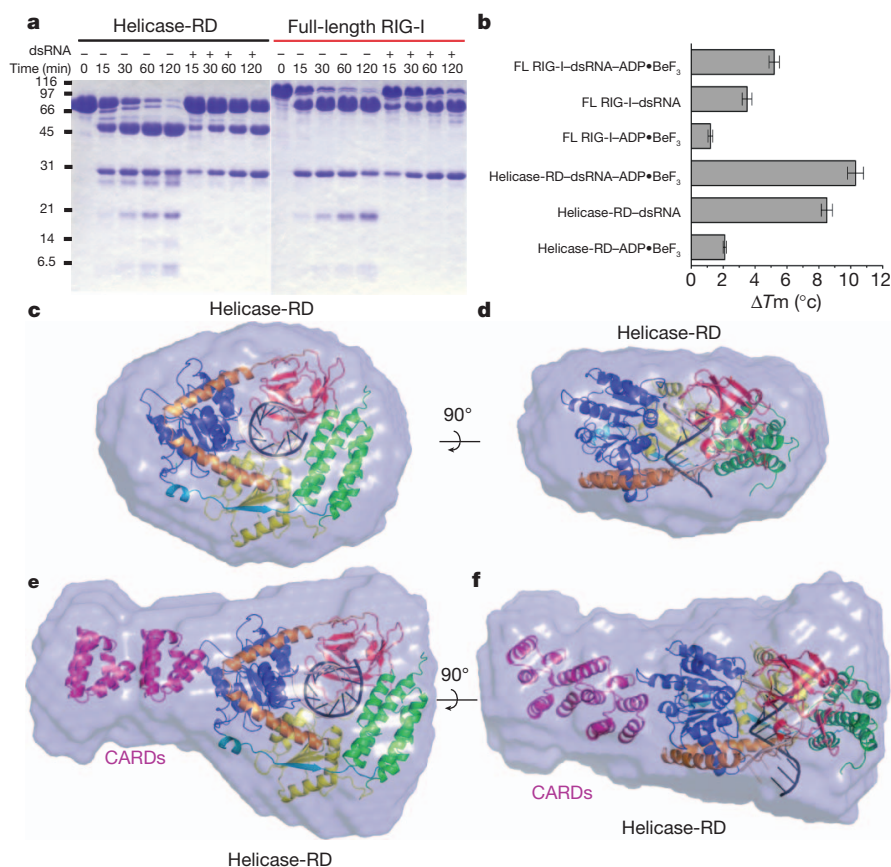


Figure 4 | Limited trypsin digestion, DSF and SAXS analyses of helicase-RD and full-length RIG-I in the presence and absence of dsRNA. **a**, SDS-PAGE analysis of a time course (minutes) of limited trypsin digestion of helicase-RD or full-length RIG-I in the absence or presence of 14 base-pair palindromic dsRNA. **b**, DSF of RIG-I helicase-RD or full-length (FL) RIG-I in the presence of 14 base-pair palindromic dsRNA and/or ADP•BeF₃ with respect to protein alone. The bar graph shows the mean melting temperature difference (ΔT_m) and the error

bars represent the standard deviation from three independent measurements. **c**, **d**, *Ab initio* envelope of helicase-RD and dsRNA overlaid with the crystal structure of helicase-RD-dsRNA (dsRNA truncated to 10 base pairs). The view in **d** is rotated 90° about a horizontal axis from panel **c**. **e**, **f**, *Ab initio* envelope of full-length RIG-I and dsRNA overlaid with the crystal structure of helicase-RD-dsRNA with two copies of CARDs added (PDB accession 2VGQ). The view in **f** is rotated 90° about a horizontal axis from panel **e**.

RIG-I. Full-length RIG-I and helicase-RD undergo conformational changes and exhibit greater stability upon RNA binding (Fig. 4). The radius of gyration (R_g) of helicase-RD and full-length RIG-I decreases upon dsRNA binding by 10 and 2.3 Å, respectively (Supplementary Table 2), consistent with the results from size-exclusion chromatography (Supplementary Fig. 1b). The SAXS Kratky plots of helicase-RD and full-length RIG-I with RNA are symmetrical parabolic curves (Supplementary Fig. 5), consistent with folded and globular complexes. The parabolic shape of the Kratky plot is lost and the peak amplitude decreases in the absence of RNA, indicative of greater flexibility¹⁹. Limited trypsin digestion and DSF shows a greater stabilization with dsRNA and a small increase in stability upon addition of the ATP analogue (Fig. 4a, b). These data indicate that RIG-I is composed of globular domains connected with flexible linkers that become ordered upon RNA binding. Such conformational stabilization upon ATP and RNA binding has been documented in other helicases^{20–22}.

A model of full-length RIG-I bound to dsRNA was established from *ab initio* SAXS envelope followed by rigid body refinement of the helicase-RD-dsRNA complex and homologous CARD structures (Supplementary Table 2 and Fig. 4c–f). The helicase-RD-dsRNA complex and two copies of the CARD are accommodated in the full-length RIG-I SAXS envelope. The two CARDs project from domain 1. Such an orientation of the CARDs would allow for interaction with downstream signalling factors^{23–28}. Interestingly, the second CARD is positioned adjacent to the V-shaped linker and close to the T770 phosphorylation site. The close proximity of the linker to the second CARD suggests a possible mechanism for the two α -helices to serve as

a hinge for RD movement upon RNA binding, leading to RIG-I activation. The model indicates that the dsRNA extends along a perpendicular axis relative to the CARDs and hence can accommodate several RIG-I molecules²⁹.

METHODS SUMMARY

Recombinant full-length RIG-I (1–925), helicase-RD (232–925), helicase (232–794), RD (795–925) and selenomethionine-derivatized helicase-RD were expressed in *Escherichia coli* and purified to homogeneity using immobilized metal ion affinity, hydroxyapatite and heparin affinity chromatography. Fluorescence anisotropy titrations were performed at 25 °C (ref. 30) ($\lambda_{\text{excitation}}$, 494 nm and $\lambda_{\text{emission}}$, 516 nm) using a fluorescein-labelled 14 base-pair dsRNA prepared by annealing 5'-GGAGAGAACCGCCU and 3'-CCUCUCUUGGCGGA-F RNA, where F is fluorescein. Crystals of the native and selenomethionine helicase-RD with palindromic dsRNA (5'-CGACGCUAGCGUCG) and ADP•BeF₃•Mg²⁺ were obtained in 25% (w/v) PEG 3350, 0.25 M NaSCN, 100 mM MOPS (pH 7.8), 3% (v/v) 2,2,2-trifluoroethanol at 20 °C by hanging drop. The crystals belong to space group P6₅22 with cell parameters $a = b = 174.9$ Å and $c = 110.9$ Å. The structure was determined by single-wavelength anomalous dispersion (SAD) to 3.2 Å resolution and refined against a 2.9 Å resolution native data set. The final model has an R_{work} and R_{free} of 0.199 and 0.287, respectively. SAXS measurements were performed on full-length RIG-I and helicase-RD in the absence and presence of dsRNA (5'-GCGCGCGCGC). Buffer subtraction and radius of gyration R_g were calculated from Guinier plots. The maximum particle size D_{max} was determined by scanning a range of values and comparing experimental scattered intensity $I(s)$ values to distance distribution function $P(r)$ transforms. Ten *ab initio* models were averaged and normalized spatial discrepancy (NSD) values were calculated. The helicase-RD-dsRNA and a homologous CARD structure were positioned into the *ab initio* model and χ^2 values were determined.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 25 July; accepted 6 September 2011.

Published online 25 September 2011.

1. Schlee, M. *et al.* Approaching the RNA ligand for RIG-I? *Immunol. Rev.* **227**, 66–74 (2009).
2. Saito, T., Owen, D. M., Jiang, F., Marcotrigiano, J. & Gale, M. Innate immunity induced by composition-dependent RIG-I recognition of hepatitis C virus RNA. *Nature* **454**, 523–527 (2008).
3. Uzri, D. & Gehrke, L. Nucleotide sequences and modifications that determine RIG-I/RNA binding and signaling activities. *J. Virol.* **83**, 4174–4184 (2009).
4. Matsumiya, T. & Stafforini, D. M. Function and regulation of retinoic acid-inducible gene-I. *Crit. Rev. Immunol.* **30**, 489–513 (2010).
5. Chattopadhyay, S. *et al.* Viral apoptosis is induced by IRF-3-mediated activation of Bax. *EMBO J.* **29**, 1762–1773 (2010).
6. Sun, Z., Ren, H., Liu, Y., Teeling, J. L. & Gu, J. Phosphorylation of RIG-I by casein kinase II inhibits its antiviral response. *J. Virol.* **85**, 1036–1047 (2011).
7. Malathi, K., Dong, B., Gale, M. Jr & Silverman, R. H. Small self-RNA generated by RNase L amplifies antiviral innate immunity. *Nature* **448**, 816–819 (2007).
8. Lu, C., Ranjith-Kumar, C. T., Hao, L., Kao, C. C. & Li, P. Crystal structure of RIG-I C-terminal domain bound to blunt-ended double-strand RNA without 5' triphosphate. *Nucleic Acids Res.* **39**, 1565–1575 (2011).
9. Lu, C. *et al.* The structural basis of 5' triphosphate double-stranded RNA recognition by RIG-I C-terminal domain. *Structure* **18**, 1032–1043 (2010).
10. Wang, Y. *et al.* Structural and functional insights into 5'-ppp RNA pattern recognition by the innate immune receptor RIG-I. *Nature Struct. Mol. Biol.* **17**, 781–787 (2010).
11. Singleton, M. R., Dillingham, M. S. & Wigley, D. B. Structure and mechanism of helicases and nucleic acid translocases. *Annu. Rev. Biochem.* **76**, 23–50 (2007).
12. Fairman-Williams, M. E., Guenther, U. P. & Jankowsky, E. SF1 and SF2 helicases: family matters. *Curr. Opin. Struct. Biol.* **20**, 313–324 (2010).
13. Dürr, H., Korner, C., Müller, M., Hickmann, V. & Hopfner, K. P. X-ray structures of the *Sulfolobus solfataricus* SWI2/SNF2 ATPase core and its complex with DNA. *Cell* **121**, 363–373 (2005).
14. Myong, S. *et al.* Cytosolic viral sensor RIG-I is a 5'-triphosphate-dependent translocase on double-stranded RNA. *Science* **323**, 1070–1074 (2009).
15. Takahashi, K. *et al.* Nonself RNA-sensing mechanism of RIG-I helicase and activation of antiviral immune responses. *Mol. Cell* **29**, 428–440 (2008).
16. Gu, M. & Rice, C. M. Three conformational snapshots of the hepatitis C virus NS3 helicase reveal a ratchet translocation mechanism. *Proc. Natl Acad. Sci. USA* **107**, 521–528 (2009).
17. Lam, A. M., Keeney, D. & Frick, D. N. Two novel conserved motifs in the hepatitis C virus NS3 protein critical for helicase action. *J. Biol. Chem.* **278**, 44514–44524 (2003).
18. Saito, T. *et al.* Regulation of innate antiviral defenses through a shared repressor domain in RIG-I and LGP2. *Proc. Natl Acad. Sci. USA* **104**, 582–587 (2007).
19. Putnam, C. D., Hammel, M., Hura, G. L. & Tainer, J. A. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* **40**, 191–285 (2007).
20. Lorsch, J. R. & Herschlag, D. The DEAD box protein eIF4A. 2. A cycle of nucleotide and RNA-dependent conformational changes. *Biochemistry* **37**, 2194–2206 (1998).
21. Polach, K. J. & Uhlenbeck, O. C. Cooperative binding of ATP and RNA substrates to the DEAD/H protein DbpA. *Biochemistry* **41**, 3693–3702 (2002).
22. Theissen, B., Karow, A. R., Kohler, J., Gubaev, A. & Klostermeier, D. Cooperative binding of ATP and RNA induces a closed conformation in a DEAD box RNA helicase. *Proc. Natl Acad. Sci. USA* **105**, 548–553 (2008).
23. Loo, Y. M. *et al.* Viral and therapeutic control of IFN- β promoter stimulator 1 during hepatitis C virus infection. *Proc. Natl Acad. Sci. USA* **103**, 6001–6006 (2006).
24. Kawai, T. *et al.* IPS-1, an adaptor triggering RIG-I- and Mda5-mediated type I interferon induction. *Nature Immunol.* **6**, 981–988 (2005).
25. Meylan, E. *et al.* Cardif is an adaptor protein in the RIG-I antiviral pathway and is targeted by hepatitis C virus. *Nature* **437**, 1167–1172 (2005).
26. Seth, R. B., Sun, L., Ea, C. K. & Chen, Z. J. Identification and characterization of MAVS, a mitochondrial antiviral signaling protein that activates NF- κ B and IRF 3. *Cell* **122**, 669–682 (2005).
27. Xu, L. G. *et al.* VISA is an adapter protein required for virus-triggered IFN- β signaling. *Mol. Cell* **19**, 727–740 (2005).
28. Gack, M. U. *et al.* TRIM25 RING-finger E3 ubiquitin ligase is essential for RIG-I-mediated antiviral activity. *Nature* **446**, 916–920 (2007).
29. Binder, M. *et al.* Molecular mechanism of signal perception and integration by the innate immune sensor retinoic acid inducible gene-I (RIG-I). *J. Biol. Chem.*, (2011).
30. Tang, G. Q., Bandwar, R. P. & Patel, S. S. Extended upstream A-T sequence increases T7 promoter strength. *J. Biol. Chem.* **280**, 40707–40713 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge access to beamlines X29 at the NSLS (National Synchrotron Light Source), LRL-CAT at APS (Advanced Photon Source), and G1 and F1 at CHESS (Cornell High Energy Synchrotron Source) and thank the NSLS, APS and CHESS staff. NSLS and APS are supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-98CH10886 and DE-AC02-06CH11357, respectively. CHESS is supported by the NSF and NIH/NIGMS through NSF award DMR-0936384, and the MacCHESS resource is supported by NIH/NCRR award RR-01646. Use of the LRL-CAT beamline facilities at Sector 31 was provided by Eli Lilly & Company. We would like to thank V. Rajagopal for initiating the biochemical experiments and guiding the project in the early stages. We thank E. Arnold, H. Berman, S. K. Burley, R. Gillilan, L. Morisco, W. Olson, T. Saito, A. Shatkin, A. Stock, H. Yang and M. Zhuravieva for providing helpful comments and assistance. This work was supported by NIH grants GM55310 to S.S.P. and AI080659 to J.M.

Author Contributions The project was initiated by M.G., J.M. and S.S.P. J.M. and S.S.P. designed and supervised the project. M.G. provided reagents and consultation. F.J. designed protein constructs and established purification protocols. A.R. generated all RNA reagents. F.J. and A.R. purified the complex and set up crystallization screens. F.J. optimized the crystal for data collection. J.M., M.T.M., F.J. and A.R. collected, processed and analysed the X-ray crystallographic data. M.T.M., F.J. and J.M. collected and analysed the SAXS data. A.R., G.-Q.T. and S.S.P. collected and analysed the RNA binding and ATPase assays. F.J. performed limited proteolysis and thermal melting assay. S.S.P. and J.M. wrote the paper and all authors contributed to editing.

Author Information Atomic coordinates have been deposited in the Protein Data Bank under accession code 3TMI. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.M. (jmarco@cabm.rutgers.edu) or S.S.P. (patelss@umdnd.edu).

METHODS

Protein expression and purification. All the protein constructs were expressed with an N-terminal 6×His-SUMO fusion. Human full-length RIG-I (1–925), helicase-RD (232–925) and helicase (232–794) were expressed in *Escherichia coli* strain Rosetta (DE3) (Novagen) as soluble proteins. The soluble fraction of helicase-RD was purified from the cell lysate using a Ni²⁺-nitrilotriacetate (Qiagen) column. The recovered protein was then digested with Ulp1 protease to remove the 6×His-SUMO tag and further purified by hydroxyapatite column (CHT-II, Bio-Rad) and heparin sepharose column (GE Healthcare). Finally, purified helicase-RD was dialysed against 50 mM HEPES (pH 7.5), 50 mM NaCl, 5 mM DTT, 10% glycerol overnight at 4 °C, snap frozen in liquid nitrogen, and stored at –80 °C. The RIG-I RD (795–925) was expressed in *E. coli* BL21 Star (DE3) cells and the soluble fraction was purified to homogeneity using a Ni²⁺-nitrilotriacetate column, cation exchange (HiTrap SP, GE Healthcare) and gel filtration chromatography. Selenomethionine (SeMet)-labelled helicase-RD was produced in Rosetta (DE3) cells grown in M9 minimal medium supplemented with 50 mg ml^{–1} L-SeMet (Sigma) and specific amino acids to inhibit endogenous methionine synthesis. The SeMet protein was purified using the same procedure as the unmodified protein.

Fluorescence anisotropy titrations. Fluorescence anisotropy measurements³¹ were carried out on a Fluoro-Max-4 spectrofluorimeter (Horiba Jobin Yvon). Fluorescein-labelled 14 base-pair dsRNA (Dharmacon) (10 nM) was titrated with full-length RIG-I, helicase domain, RD or helicase-RD in 50 mM Tris-acetate (pH 7.5), 100 mM Na-acetate, 10 mM Mg-acetate, 5% glycerol, 5 mM DTT, 0.05% Tween 20 buffer. 5'-ppp dsRNA (GGAGAGAACCGCCU) was transcribed using T7 RNA polymerase, PAGE purified, and annealed to a complementary, fluorescein-labelled RNA. Fluorescein anisotropy was measured at 25 °C with excitation at 494 nm and emission at 516 nm. The helicase-RD titrations were also performed at lower dsRNA concentrations (1 and 2 nM) to obtain an accurate K_d . The observed fluorescence anisotropy (r_{obs}) as a function of protein concentration (P_t) was fit to equations (1) and (2) to obtain the equilibrium dissociation constant, K_d .

$$r_{\text{obs}} = r_{\text{fb}} + r_{\text{f}}(1 - f_{\text{b}}) \quad (1)$$

where r_{f} and r_{b} are the anisotropy values of free RNA and of the complex, f_{b} is the fraction of RNA bound in the protein–RNA complex and $f_{\text{b}} = [PR]/[R_t]$ (PR is the concentration of the protein–RNA complex and R_t is the total RNA concentration).

$$[PR] = \frac{(K_d + [P_t] + [R_t]) - \sqrt{(K_d + [P_t] + [R_t])^2 - 4[P_t][R_t]}}{2} \quad (2)$$

Initial anisotropy of the free fluorescein-labelled dsRNA was 0.10 for all experiments. The measurements and errors are from two independent experiments in Table 1.

ATPase activity. A time course (0–30 min) of the ATPase reaction was performed using RIG-I helicase-RD (5 nM), ATP (1 mM) spiked with [γ -³²P]ATP with or without 14 base-pair dsRNA (80 nM) in buffer containing 50 mM MOPS (pH 7.4), 5 mM MgCl₂, 5 mM DTT, 0.01% Tween 20 at 37 °C. The ATPase activity of the RIG-I helicase domain was measured using a higher amount of protein (100 nM) and RNA (1 μM). The quenched reactions were analysed by PEI-Cellulose-F TLC (Merck) developed in 0.4 M potassium phosphate buffer (pH 3.4). The TLC plates were exposed to a phosphorimager plate, analysed on a Typhoon phosphor-imager, and quantified using ImageQuant software. The ATPase rate was determined from the plots of [Pi] produced versus time and the rate constant values were calculated by dividing the ATPase rate by the respective enzyme concentration. The mean rate constant from two independent measurements and range is shown in Table 1.

Analytical size-exclusion chromatography. Analytical size-exclusion chromatography was carried out on an AKTA FPLC system (GE Healthcare). Proteins were loaded on to Superdex 200 10/300 GL column (GE Healthcare) equilibrated with a buffer containing 50 mM HEPES (pH 7.5), 50 mM NaCl, 5 mM MgCl₂, 5 mM DTT. The eluate was monitored by ultraviolet absorbance at 280 nm. For the complex, helicase-RD was incubated with 14 base-pair dsRNA at the molar ratio of 1:1.2 on ice for 15 min before the sample was applied onto the column.

Preparation of helicase-RD–dsRNA. The palindromic RNA oligonucleotide (5'-CGACGCUAGCGUCG-3') (Dharmacon) was deprotected and desalted into 20 mM potassium phosphate (pH 7.4) before use. The dsRNA was prepared by incubating the RNA at 95 °C for 1 min followed by gradual cooling to 4 °C. The resulting dsRNAs were mixed with purified helicase-RD in a RNA:protein molar ratio of 1.2:1, incubated at room temperature (20 °C) for 15 min, and then purified by size-exclusion chromatography (Superdex200, GE Healthcare) with

an elution buffer of 50 mM HEPES (pH 7.5), 50 mM NaCl, 5 mM DTT, 5 mM MgCl₂.

Crystallization and X-ray diffraction data collection. Helicase-RD–dsRNA–ADP•BeF₃ ternary complex was reconstituted by incubating 0.17 mM helicase-RD with 0.17 mM dsRNA, 2 mM ADP, 2 mM BeCl₂ and 10 mM NaF on ice for 30 min before crystallization. Crystals of native and SeMet-substituted complex were grown by the hanging-drop vapour diffusion method at 20 °C. Aliquots (2.5 μl) of 15 mg ml^{–1} of helicase-RD–dsRNA–ADP•BeF₃ complex in 50 mM HEPES (pH 7.5), 50 mM NaCl, 5 mM DTT, 5 mM MgCl₂ were mixed with 2.5 μl of reservoir solution containing 25% (w/v) PEG 3350, 0.25 M NaSCN, 100 mM MOPS (pH 7.8), 3% (v/v) 2,2,2-trifluoroethanol. Crystals appeared after 2–3 days, and they grew to a maximum size of 0.15 × 0.15 × 0.5 mm over the course of 8 days. 5-iodo-uridine derivative helicase-RD–dsRNA–ADP•BeF₃ crystals were grown under identical conditions. For cryogenic data collection, crystals were transferred into crystallization solutions containing 5% (v/v) (2R,3R)-(-)-2,3-butanediol as cryoprotectant and then flash-cooled at 100 K. SeMet SAD data set was collected at the X29A beamline of the NSLS. Native data set was recorded at the LRL-CAT 31-ID beamline of the APS. Data from iodouridine derivative crystals were collected at the CHESS F1 beamline. All diffraction data were integrated using iMosflm and scaled in SCALA³².

Structure determination and refinement. Phases were determined using the SAD method. SHELXD/HKL2MAP³³ detected a total of 12 out of 14 possible selenium sites in the asymmetric unit. Initial phases were calculated with Phaser³² and followed by density modification by DM³². Phase extension to 3.2 Å by SOLOMON³² produced an electron-density map into which most of the protein and RNA residues could be built unambiguously. The model was built using Coot³⁴ and refined in PHENIX³⁵. The final model, comprising ADP•BeF₃, 12 base-pair dsRNA, RIG-I residues 240–495, 504–522, 527–687 and 690–923 has R_{work} and R_{free} values of 0.199 and 0.287, respectively. Model validation demonstrated no outliers and 90% of the residues located the most favourable region of the Ramachandran plot³⁵. Statistics of the data processing and structure refinement are summarized in Supplementary Table 1.

SAXS and structural modelling. SAXS data were collected at the CHESS beamline G1 using a Finger Lakes CCD X-ray detector with a sample-to-detector distance of 1,450 mm, covering the range of scattering vectors $0.01 < s < 0.233 \text{ \AA}^{-1}$, where $s = 4\pi \sin \theta / \lambda$ (2θ is the scattering angle and $\lambda = 1.296 \text{ \AA}$). All samples were buffer exchanged into 50 mM HEPES (pH 7.5), 50 mM NaCl, 5 mM DTT, 5% glycerol by size-exclusion chromatography (Hiload 16/26 Superdex200, GE Healthcare) to minimize the discrepancies in background subtraction. SAXS data were reduced using Data Squeeze. Various programs in ATSAS software package were used to process and evaluate the scattering data. Radius of gyration (R_g) was analysed using the Guinier approximation with low angle data ($s < 1.3/R_g$) using PRIMUS³⁶. The residuals from the Guinier plots did not show signs of protein aggregation. The probability distribution of distances between scattering atoms within the macromolecule, $P(r)$, and the maximum atom pair distance, D_{max} , were determined from the scattering data using GNOM³⁷. The program DAMMIF³⁸ was used to calculate low-resolution *ab initio* reconstructions from experimental SAXS profiles. Ten to twenty independent models were aligned, filtered and averaged based on the occupancy using SUPCOMB and DAMAVER³⁹ to reconstruct the final *ab initio* envelope, as judged by averaged normalized spatial discrepancies (NSD) less than 1.0. The structure modelling was refined against the solution scattering data by rigid body docking using SASREF program⁴⁰. The orientation and position of individual domains in the structure models were further manually adjusted to minimize the discrepancies (χ^2) between the calculated scattering intensities and experimental scattering intensities computed using program CRY SOL⁴¹. Kratky plots were calculated using Origin (OriginLab)³⁶.

DSF. Thermal shift assay was conducted with 10 μM of RIG-I helicase-RD or full-length RIG-I with or without 12 μM of 14 base-pair palindromic dsRNA and/or 2 mM ADP•BeF₃ in 50 mM HEPES (pH 7.5), 50 mM NaCl, 5 mM MgCl₂, 5 mM DTT and a 5× dilution of SYPRO Orange dye (Invitrogen) as described⁴². The fluorescence signal as a function of temperature was recorded using a Real Time PCR machine (Applied Biosystems). The temperature gradient is performed in the range of 25–80 °C with a ramp of 0.2 °C over the course of 60 min. Control assays were carried out with buffer in the presence or absence of RNA. Data were analysed with the Excel-based worksheet DSF analysis, and the Boltzmann model was used to fit the fluorescence data to obtain the midpoint temperature for the thermal protein unfolding transition (T_m) using the curve-fitting software Prism.

Limited trypsin proteolysis. Limited proteolysis with trypsin (Roche) was performed using 120 μg of purified full-length RIG-I or helicase-RD in the absence or presence of palindromic dsRNA and incubated with trypsin at a protein:protease mass ratio of 300:1. The reaction mixtures were maintained at room temperature

and aliquots were removed at 15, 30, 60 and 120 min. The reaction was stopped by the addition of SDS-PAGE loading buffer and analysed by SDS-PAGE.

31. Tang, G. Q., Bandwar, R. P. & Patel, S. S. Extended upstream A-T sequence increases T7 promoter strength. *J. Biol. Chem.* **280**, 40707–40713 (2005).
32. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242 (2011).
33. Schneider, T. R. & Sheldrick, G. M. Substructure solution with SHELXD. *Acta Crystallogr. D* **58**, 1772–1779 (2002).
34. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
35. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
36. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. & Svergun, D. I. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Cryst.* **36**, 1277–1282 (2003).
37. Semenyuk, A. V. & Svergun, D. I. GNOM: a program package for small-angle scattering data processing. *J. Appl. Cryst.* **24**, 537–540 (1991).
38. Svergun, D. I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76**, 2879–2886 (1999).
39. Volkov, V. V. & Svergun, D. I. Uniqueness of ab initio shape determination in small-angle scattering. *J. Appl. Cryst.* **36**, 860–864 (2003).
40. Petoukhov, M. V. & Svergun, D. I. Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys. J.* **89**, 1237–1250 (2005).
41. Bernadó, P., Mylonas, E., Petoukhov, M. V., Blackledge, M. & Svergun, D. I. Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.* **129**, 5656–5664 (2007).
42. Niesen, F. H., Berglund, H. & Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature Protocols* **2**, 2212–2221 (2007).

EXTRASOLAR PLANETS

Homing in on another Earth

The identification of the closest analogue of Earth so far, orbiting another star, suggests that small planets are common, and that the discovery of a candidate habitable planet in an alien star system could be just around the corner.

JACOB BEAN

In the hunt for planets around stars other than the Sun, astronomers' primary objective is to find a planet that is teeming with life. A milestone on the path to this goal is the discovery of Earth-sized planets orbiting their parent stars in the 'habitable zone' — the range of distances from the star at which the temperature would be just right for liquid water to be present on a planet's surface. But which stars harbour such planets, how common are they, and what are their basic characteristics? Although astronomers can't yet answer these questions, a paper to be published in *Astronomy & Astrophysics* by Pepe *et al.*¹ presents the discovery of several planets that marks a significant step towards changing this impasse.

Pepe and colleagues detected five small planets orbiting parent stars that are slightly smaller and cooler than the Sun. One of the planets is only 3.6 times the mass of Earth and is in an orbit that teases the inner edge of its host star's habitable zone. This is the closest that astronomers have yet come to finding another Earth. Furthermore, the relative ease with which these and other previously reported small planets have been found by the same group implies that the frequency of such planets around Sun-like stars is on the order of tens of per cent.

The authors made their new discoveries¹ using the radial-velocity technique, which is an indirect method based on measuring the periodic change in speed of a star caused by the gravitational tug of orbiting bodies. Until this year, most planet discoveries were made with this method. But there has been a widely held opinion that the radial-velocity technique will not be able to find candidate habitable Earth-mass planets despite its success at finding Jupiter-mass planets. The reasoning behind this is that the variability of the visible surfaces of stars, as a result of magnetic activity, pulsations and the turbulence of the plasma in their atmospheres, creates noise in radial-velocity measurements, and this noise is larger than the signal induced by an Earth-sized planet in the habitable zone. Also, the technical challenge of building instruments sensitive enough to detect the subtle indications of such planets, even around perfectly 'quiet' stars, was considered too challenging.

Pepe and colleagues' detections¹ were

enabled by a combination of exquisite instrumentation, painstaking analysis and an ingenious observational technique. Using an ultra-stable instrument on a dedicated telescope, and with the benefit of ten years' experience in refining their calibration strategy, the authors set out to intensively monitor ten of the quietest known nearby stars. Their monitoring campaign involved making multiple measurements per night to average over the stellar noise cycles that had plagued previous studies. Using this approach, they have broken through the previously inviolate one-metre-per-second radial-velocity barrier and detected planets that have signals less than tenfold larger than that of an Earth analogue,

One of the planets is only 3.6 times the mass of Earth and is in an orbit that teases the inner edge of its host star's 'habitable' zone.

for which the signal is 9 cm s^{-1} . Pepe *et al.* were also careful to show that the detected signals were most probably attributable to orbiting planets rather than to intrinsic variations in the stars themselves. They did this by demonstrating that the orbital periods determined for the planets were distinct from the host stars' rotation periods, and also that there were no correlations between the detected signals and the diagnostics of stellar activity.

Despite being a clear breakthrough, there are some limitations to Pepe and colleagues' study. One limitation is that the newly detected planets are poorly characterized. We do not know the compositions of the planets because the radial-velocity method yields no information about the densities of the planets it detects. Also, we have little knowledge of how elliptical the planets' orbits are (their orbital eccentricity), because the detected signals are relatively small. Both composition and orbital eccentricity are crucial parameters for assessing a planet's habitability.

Furthermore, although Pepe and colleagues' results¹, and those described in another paper from the same group², hint at a large population of small planets orbiting Sun-like stars, the statistics are weak because these planets are difficult to detect and therefore the samples are

incomplete by an unknown amount. Identifying a large sample of similar planets, and studying them by different methods, would advance the field. However, Pepe *et al.* show that the full potential of the radial-velocity technique has not yet been realized. This work warrants the building of a generation of radial-velocity instruments that are even more stable than those now in use, so that instrumental noise is no longer a barrier to the discovery of Earth-like planets with this method.

NASA's ongoing Kepler mission, which finds planets using the transit technique, announced a haul of more than 1,000 new planet candidates earlier this year, and is on track to identify habitable-zone planets³. Because the transit technique detects planets by measuring periodic decreases in a star's brightness when a planet passes in front of it, the issue of stellar variability is also a limiting factor for this approach, and early results indicate that the stars Kepler is looking at are significantly more variable than expected⁴. This means that an extension beyond the nominal 3.5-year mission duration might be necessary for Kepler to securely detect Earth-sized planets in the habitable zones of Sun-like stars.

In the long run, astronomers aim to study the atmospheres of the small worlds revealed by radial-velocity and transit surveys to obtain further insight into the planets' habitability and even, perhaps, their state of inhabitation. The investigations that will be enabled by NASA's planned James Webb Space Telescope, which is currently at risk of cancellation, are the cornerstone of astronomers' next plans in this area. Nevertheless, it is clear that the search for other Earths is gathering pace. The exciting results from Kepler, and the remarkable advances in the radial-velocity technique demonstrated by Pepe *et al.*, show that the race is well and truly on. ■

Jacob Bean is at the Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA. e-mail: jbean@cfa.harvard.edu

1. Pepe, F. *et al.* *Astron. Astrophys.* (in the press); preprint at <http://arxiv.org/abs/1108.3447> (2011).
2. Mayor, M. *et al.* preprint at <http://arxiv.org/abs/1109.2497> (2011).
3. Borucki, W. J. *et al.* *Astrophys. J.* **736**, 19 (2011).
4. Gilliland, R. L. *et al.* *Astrophys. J.* (in the press); preprint at <http://arxiv.org/abs/1107.5207> (2011).

Can algal uptake stop NO_3^- pollution?

ARISING FROM Cardinale, B. J. *Nature* **472**, 86–89 (2011)

The influence of biodiversity on ecosystem function has been of interest to community ecologists for decades. Recently, Cardinale¹ reported that biodiversity affects nitrate (NO_3^-) uptake in algal communities and that, as a result, biodiversity may help mitigate nutrient pollution. Although Cardinale's conclusions about niche partitioning are interesting (figure 2 in ref. 1), his extension of these findings to problems of nutrient pollution is premature. Algal uptake is only a short-term nitrogen sink; control of NO_3^- pollution requires long-term solutions.

Nitrate removal in streams typically occurs via two pathways. Denitrification is the microbial conversion of NO_3^- to N_2 gas (Fig. 1). This is the most important means of permanent NO_3^- removal from streams. However, it is not considered in Cardinale's study.

The second NO_3^- uptake pathway is assimilation by primary producers (in this case, algae). Algae acquire nutrients to fulfil physiological demands and support growth, but these same nutrients are later released after cell death or disruption (such as the brushing treatment applied to some flumes to create heterogeneity). Algal uptake leads to temporary storage of nitrogen in algal tissues (turn-over time ranges from ~17.5–47 days^{2–4}), and can have significant short-term effects on nitrogen export to downstream ecosystems. But, because benthic algae do not accumulate over long (annual or inter-annual) periods, this uptake mechanism represents only a transient storage pool and thus has no enduring effect on downstream water quality (Fig. 1).

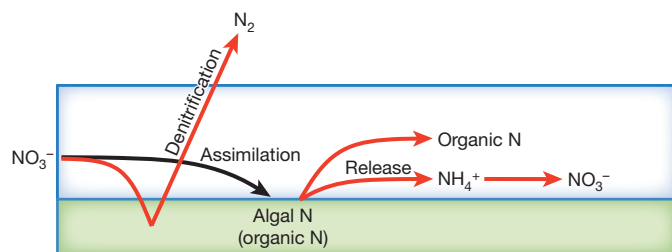


Figure 1 | Nitrogen cycling in streams. Red arrows indicate processes not measured by Cardinale¹. Storage in groundwater and sediments, N_2 fixation and fluxes to other ecosystems (for example, floodplains) may also occur.

Algae could potentially mediate permanent nitrogen removal in streams by facilitating increased rates of denitrification^{5,6}. Indeed, Cardinale's results raise the intriguing possibility that algal diversity could affect denitrification rates via changes in the availability of nitrogen, carbon or oxygen to denitrifiers. However, until this potential effect is demonstrated, conclusions about biodiversity effects on the control of aquatic nitrogen pollution are premature.

We agree with Cardinale that nitrogen enrichment threatens the integrity of aquatic ecosystems and that nitrogen removal is a critical ecosystem service provided by streams⁷. And, although Cardinale has elucidated a potential mechanism for biodiversity to influence stream nutrient cycling, for this work to be extended to pollution control in natural ecosystems it must be tested in environments incorporating critical biogeochemical pathways—namely, denitrification and nitrogen fixation—and it must consider long-term nitrogen dynamics.

Helen M. Baulch¹, Emily H. Stanley¹ & Emily S. Bernhardt²

¹Center for Limnology, University of Wisconsin, Madison, Wisconsin 53706, USA.

e-mail: helenmbaulch@gmail.com

²Biology Department, Duke University, Durham, North Carolina 27708, USA.

1. Cardinale, B. J. Biodiversity improves water quality through niche partitioning. *Nature* **472**, 86–89 (2011).
2. Merriam, J. L. *et al.* Characterizing nitrogen dynamics, retention and transport in a tropical rainforest stream using an *in situ* ^{15}N addition. *Freshw. Biol.* **47**, 143–160 (2002).
3. Mulholland, P. J. *et al.* Nitrogen cycling in a forest stream determined by a ^{15}N tracer addition. *Ecol. Monogr.* **70**, 471–493 (2000).
4. Tank, J. L. *et al.* Analysis of nitrogen cycling in a forest stream during autumn using a ^{15}N -tracer addition. *Limnol. Oceanogr.* **45**, 1013–1029 (2000).
5. Ishida, C. K., Arnon, S., Peterson, C. G., Kelly, J. J. & Gray, K. A. Influence of algal community structure on denitrification rates in periphyton cultivated on artificial substrata. *Microb. Ecol.* **56**, 140–152 (2008).
6. McMillan, S. K., Piehler, M. F., Thompson, S. P. & Paerl, H. W. Denitrification of nitrogen released from senescing algal biomass in coastal agricultural headwater streams. *J. Environ. Qual.* **39**, 274–281 (2010).
7. Mulholland, P. J. *et al.* Stream denitrification across biomes and its response to anthropogenic nitrate loading. *Nature* **452**, 202–205 (2008).

Author Contributions All authors contributed to writing and revising this paper.

Competing financial interests: declared none.

doi:10.1038/nature10418

Cardinale reply

REPLYING TO H. M. Baulch, E. H. Stanley & E. S. Bernhardt *Nature* **477**, doi:10.1038/nature10418 (2011)

Over the last 20 years, 59 experiments have quantified how the richness of plants and algae influence concentrations of inorganic nitrogen in soil or water¹. Of these, 86% have shown that the concentration of nitrogen decreases as biodiversity increases—by an average of 48%. The primary contribution of my study² was to identify a biological mechanism that is likely to explain these biodiversity effects. Using stream mesocosms, I showed that the impacts of algal diversity on nitrogen dynamics are controlled by niche partitioning—a long presumed, but rarely demonstrated mechanism. Baulch, Stanley and

Bernhardt³ have questioned whether my findings have any implications for managing water quality in 'real' streams, as I suggested. They argue that nitrogen assimilation by algae cannot influence long-term nitrogen retention due to high turnover of algal biomass and rapid recycling of nitrogen, and they suggest that the only permanent loss of nitrogen from a stream is via denitrification.

Baulch, Stanley and Bernhardt³ are correct in saying that my study does not reveal the ultimate, long-term fate of nitrogen in streams. Although assimilation represents the critical first step in nitrogen

removal from water, it must be followed by some other process that leads to long-term storage or permanent removal from a stream if there is to be any meaningful improvement in water quality. Increased rates of assimilation very likely increase the probability that other long-term mechanisms operate, but my study only explicitly linked diversity to assimilation. So I agree that researchers now need to take the next step and examine the impacts of biodiversity on additional pathways involved in the nitrogen cycle.

But as we take these next steps, I worry that Baulch, Stanley and Bernhardt³ have given an incomplete roadmap. From their comment, one might get the impression that everyone agrees on what processes control the fate of nitrogen in streams, and that denitrification is the sole process we need to focus on. I do not believe this to be the case.

Baulch, Stanley and Bernhardt³ present what I call the 'quasi-chemostat' model of a stream, which assumes that streams are in a steady state such that there is no net accrual of biomass, and no net storage of material. As nutrients flow into a stream, they are assimilated and remineralized at comparable rates by organisms with high turnover. The nutrients are not transformed in any way, such that nitrogen simply 'spirals' downstream in the same biologically active form. They argue³ that the one exception to these chemostat-like assumptions is that denitrification may transfer some nitrogen from the stream to air.

Many researchers, including myself, think that the quasi-chemostat model makes overly simplistic assumptions that are not supported by existing data. Dozens of ¹⁵N tracer studies have shown the following. First, assimilation accounts for an average of 84% of NO₃⁻ removal from stream water^{4,5}, whereas denitrification accounts for just 16%⁶. Second, up to 67% of assimilated nitrogen can go into primary consumers⁷ (a pathway that Baulch, Stanley and Bernhardt³ do not consider), much of which can be lost to terrestrial environments via emergence or consumption by terrestrial predators^{8,9}. Third, some studies indicate that 25% or more of imported nitrogen can end up in depositional habitats¹⁰ or hyporheic zones¹¹ where it has residence times on the scale of years^{12,13}. Last, several studies show that the vast majority of nitrogen that enters a stream reach is transformed and exported downstream in a particulate form that is less available for biological uptake^{7,12}.

These findings all violate the assumptions that Baulch, Stanley and Bernhardt³ have made about nitrogen cycling in streams. So although they make a useful point that we need to examine how biodiversity influences the suite of processes that influence the long-term fate of assimilated nitrogen, I believe that we would benefit from taking a broader view of the nitrogen cycle than depicted in figure 1 of Baulch, Stanley and Bernhardt³.

Bradley J. Cardinale¹

¹University of Michigan, School of Natural Resources & Environment, 440 Church Street, Ann Arbor, Michigan 48109-1041, USA.
e-mail: bradcard@umich.edu

1. Cardinale, B. J. *et al.* The functional role of producer diversity in ecosystems. *Am. J. Bot.* **98**, 572–592 (2011).
2. Cardinale, B. J. Biodiversity improves water quality through niche partitioning. *Nature* **472**, 86–89 (2011).
3. Baulch, H. M., Stanley, E. H. & Bernhardt, E. S. Can algal uptake stop NO₃⁻ pollution? *Nature* **477**, doi:10.1038/nature10418 (2011).
4. Hall, R. O. *et al.* Nitrate removal in stream ecosystems measured by ¹⁵N addition experiments: total uptake. *Limnol. Oceanogr.* **54**, 653–665 (2009).
5. Mulholland, P. J. *et al.* Nitrate removal in stream ecosystems measured by ¹⁵N addition experiments: denitrification. *Limnol. Oceanogr.* **54**, 666–680 (2009).
6. Mulholland, P. J. *et al.* Stream denitrification across biomes and its response to anthropogenic nitrate loading. *Nature* **452**, 202–205 (2008).
7. Dodds, W. K. *et al.* Quantification of the nitrogen cycle in a prairie stream. *Ecosystems* **3**, 574–589 (2000).
8. Sanzone, D. M. *et al.* Carbon and nitrogen transfer from a desert stream to riparian predators. *Oecologia* **134**, 238–250 (2003).
9. Nakano, S. & Murakami, M. Reciprocal subsidies: dynamic interdependence between terrestrial and aquatic food webs. *Proc. Natl Acad. Sci. USA* **98**, 166–170 (2001).
10. Bernot, M. J. & Dodds, W. K. Nitrogen retention, removal, and saturation in lotic ecosystems. *Ecosystems* **8**, 442–453 (2005).
11. Triska, F. J., Kennedy, V. C., Avanzino, R. J., Zellweger, G. W. & Bencala, K. E. Retention and transport of nutrients in a third-order stream in northwestern California: hyporheic processes. *Ecology* **70**, 1893–1905 (1989).
12. Hall, R. O., Baker, M. A., Arp, C. D. & Koch, B. J. Hydrologic control of nitrogen removal, storage, and export in a mountain stream. *Limnol. Oceanogr.* **54**, 2128–2142 (2009).
13. Ashkenas, L. R., Johnson, S. L., Gregory, S. V., Tank, J. L. & Wollheim, W. M. A stable isotope tracer study of nitrogen uptake and transformation in an old-growth forest stream. *Ecology* **85**, 1725–1739 (2004).

doi:10.1038/nature10419